

Britain hazards embryo research

The British government's much-delayed response to the Warnock Committee's report on the manipulation of human embryos would unwisely let parliament decide whether research is to be allowed.

THAT there should be restraints of some kind on the manipulation of human embryos has been clear for at least the past two decades. In the past five years, with the development of techniques for transferring artificially fertilized ova to human uteri, the urgent need has been not merely to draw a line between the permissible and the forbidden, but also to devise a mechanism for administering the legal framework of restraint. Given the contributions of British scientists to the practice of *in vitro* fertilization, the British government inevitably finds itself among the first in the legislative field, after the state of Victoria in Australia. Does that explain why it has taken more than three years for the government to describe in general terms the legislation it plans for the permanent regulation of the manipulation of human embryos, both in the treatment of infertility and otherwise, in research for example (see p. 409)?

The delay since the publication of the Warnock report, which the government's proposals broadly follow, has been particularly intolerable to practitioners and to the members of the Voluntary Licensing Authority, set up by the Medical Research Council and the Royal College of Obstetrics and Gynaecology, who have discharged with ingenuity and distinction the thankless task of regulating without statutory authority. It will seem to them a poor reward that the government has decided not to make up its mind on the question whether research with human embryos will be allowable under any circumstances, but instead to take pot luck with a free vote in the House of Commons.

That does not, of course, imply that members of parliament cannot be trusted to make rational decisions. Rather, it is the case that the coming year is likely to be over-charged with controversy over the rights and wrongs of British abortion law, chiefly because Mr David Alton, a Liberal member of the House of Commons, has introduced a bill to reduce the legal limit on the gestation time at which fetuses may be legally aborted (which requires two physicians to certify that the putative mother's physical or mental health would otherwise be impaired). Mr Alton proposes that the legal limit should be reduced from 28 weeks to 18 weeks, which is probably sensible; there is room for argument about the numbers, but there is no doubt that the technology for the preservation of fetuses born prematurely has improved in the past 25 years, while there have also been many improvements of the diagnosis of pregnancy, and changing manners have diminished the likelihood that pregnant women will for long be innocent of their condition.

Trouble

The trouble is that Alton's innocuous bill is certain also to provide an opportunity for those who hold that abortion should not be permitted in any circumstances to re-run the arguments that echoed round the British parliament in the 1960s. It is inevitable that this old controversy will inflame the argument about embryo research that will follow. Indeed, it is also likely that the prospect of legislation on embryo research will confuse the issue about the abortion law; there are many people in the House of Commons eager for a chance to rattle the threat that devilish experiments with human embryos are around the corner to support the case that abortion should be forbidden. That is

why it would have been much better if the government had put its weight behind a proposal that embryo research should be allowed under certain conditions.

Why is that important? The simple reason is that there is much of potential benefit to be learned not merely about the development of the human embryo, and the reasons why the course of development is not always within the norm, from carefully regulated embryo research. The causes of spina bifida and other congenital malformations of the spinal tract, for example, are likely to be accessible to studies of human embryos in the few weeks after fertilization. The better understanding of the sequence of development of human organs would be valued, not merely for its own sake but as a foundation for later studies of teratogenesis and natural uterine influences on development.

Research

There is also a sense in which the government's proposals entail that there should be at least some research with embryos; the draft clause the government plans to introduce if parliament decides to prohibit research would forbid all procedures with embryos other than those intended to assist their transfer to a uterus or "to ascertain the suitability of that embryo for the intended transfer". The genetic analysis of early embryos would be a natural goal, but one that will not be attainable without further research, which will no doubt be carried through somehow, somewhere and at some time. So the British government seems to be proposing that, if parliament decides that research will not be allowed under any circumstances, British practitioners will have to rely for the techniques for ascertaining "suitability" on work carried out elsewhere. Does that make sense?

Why all the fuss about embryo research? The government's white paper notes (in an appendix) the sharp difference of opinion between those who respectively allow and disallow embryo research that the former "would not generally accept the proposition that the human embryo should, from the point of conception, be regarded as having the same full human status as a child". The latter, the white paper says, hold that the eventual destruction of embryos, a necessary consequence of embryo research, "is tantamount to murder". This apposition can only sharpen the government's own difficulties in the months ahead.

The truth is that both sides in the argument agree that embryo research must at the very least be regulated because human embryos are potentially human beings, and are deserving of respect on that account. And while a one-day embryo cannot hope to survive without the benefit of a human uterus, there are many among the supporters of embryo research who would agree that the status of such an embryo is not very different from that of, say, an 18-week fetus; both are potentially human beings. The case for a continuation of research with human embryos does not rest on the assertion that early embryos are otherwise, but on the belief that the social benefits of continuing research would be considerable and the conviction that the manipulation of small clumps of cells can be done without raising in the minds of practitioners the sense of having been killing people. Ironically, abortion (also justified largely on social grounds) is more evidently offensive. □



Togetherness by debate

Public debate about SDI (and AIDS) has made few converts but may have bridged a gap.

If the success of a contribution to a debate is measured by the number of people whose minds are changed, this year's report by the American Physical Society on the US Strategic Defense Initiative (SDI) is a failure. To some, it demonstrates how technically far-fetched are the goals of SDI, but the authors of the report are careful not to say that any particular ingredient of SDI is a technical impossibility, which has been construed by supporters of SDI as a guide to the areas and amount of research needed, and thus as a blueprint for future spending. Privately, SDI supporters are scornful, and say that the report tells them nothing new. Publicly, they poke fun at its authors and remind them of eminent scientists who declared that powered flight, or nuclear power, or space travel are impossible.

The US Department of Defense also appears to relish tying up reports in extended and often seemingly perverse classification reviews — on occasion going so far as to classify sections of a report containing little more than quoted testimony from open Congressional hearings — and, after declaring them free from information that might jeopardize national security, pouncing on them for not being sufficiently up-to-date. Such was the fate of the APS report, and a similar fate seems to await an extensive analysis of SDI by the Congressional Office of Technology Assessment, now stalled in the Pentagon.

The technical debate, whether SDI can work, and the political debate, whether it should be attempted, apparently lead independent lives. Even a discussion between Kumar Patel, chairman of the APS panel that produced the report, and Louis Marquet, of the SDI Organization, before an audience of physicists, soon turned ideological, with Marquet responding to Patel's criticisms by declaring that strategic defense is a noble aim and more research must be done. This state of affairs no doubt distresses those who think politicians should decide such matters from a base of technological competence. But how well informed must politicians be before they can make an informed decision? And even if those who control funds knew and understood as much about SDI as the authors of the APS report, would their decision-making be any easier, or receive more universal acclaim? One presumes that scientists on both sides of the debate are working with the same laws of physics, and the evident fact that they arrive at different conclusions hardly encourages the view that scientifically knowledgeable politicians would somehow do better, and reach an apolitical consensus.

But professional people should not be too depressed. Their contributions to the SDI debate may not have caused anyone to switch sides, but they have accomplished something more subtle: they have moved the sides closer together, even though the participants may be reluctant to admit it. Supporters of SDI, rather than admitting that President Reagan's vision of an impenetrable shield to render missiles obsolete is no more than wishful thinking, now declare that of course they never saw that ideal as an immediate aim but as a long-term theme behind continued research: 'enhanced deterrence' is their watchword now. Opponents, on the other hand, having begun by roundly declaring SDI both a physical impossibility and a destabilizer of the strategic balance, have difficulty bringing themselves to admit that defence as such is not evil, and that some research into defensive strategies is an obvious and prudent course.

What happens now is easy to predict. Everyone agrees that defensive measures are useful and that research must be done. Probably sooner rather than later, the SDI Organization will be formally disbanded, but its objectives and funds will find shelter under different umbrellas in the Department of Defense. Some research areas will be abandoned, and new ones will be added. Given the amount of money being spent, some working devices

will be built, although they will be less efficient or useful than their inventors might hope. And the arms race will continue.

Although the nature of the debate is quite different, ideology also plays a large role in decisions about how best to cope with the spread of AIDS, about which there has been more willingness to conduct a rational debate. Perhaps this is because ideologically charged phrases such as "national security" are not involved, although certainly reference to sexual practices can change a discussion into a shouting match. Or perhaps there is greater confidence in the track record of researchers to curb potentially fatal epidemics than there is for physicists to prevent nuclear war. Whatever the reason, reports by Surgeon General C. Everett Koop and the Institute of Medicine have been well received, and their recommendations widely adopted.

The Institute of Medicine says it will now update its successful report *Confronting AIDS*. The nature of the AIDS problem and the national and international response to it have been changing at a rate that a re-evaluation after only two years is entirely appropriate. But the decision to proceed with a second edition signals the level of commitment that the institute has to its recommendations. It would be easy enough to grow discouraged that some of the most pressing and urgent problems identified in the original edition have still not been tackled. There is still no body identified to coordinate the US response to AIDS, no strong leadership from Reagan on the subject and no large-scale public education campaign. But the report has forced the Reagan Administration to address these issues in a formal sense, and that is all to the good. Again, it seems, even when a public argument changes few people's minds overtly, the argument creates a vocabulary in which the two sides can talk to each other. □

Distracting occasions

This year's solstice season seems to be more than usually provided with a portentous calendar.

THIS could be a momentous week in the history of the world. Mr Mikhail Gorbachev is travelling to Washington to sign a treaty with President Reagan that will abolish Euromissiles; quite apart from its intrinsic importance, the treaty will break new ground in arms-control arrangements by its detailed prescription of arrangements for verification of compliance by on-site inspection at the volition of the inspecting side. Mr Reagan will be able to make a stirring speech and put the onus of ratification on the US Senate which, when the chips are down, dares not say no.

Much the same is happening with the US dollar, which is now subsiding quietly but steadily in the wake of the long haggle between the White House and the US Congress in their nearly abortive attempt to cut at least \$23,000 million from the current US budget deficit. In the short term, the value of a currency on the foreign exchanges is determined by the supply of and the demand for it; the US dollar has been falling against other currencies because creditors outside the United States have turned shy of lending more. Indeed, the great stock-market crash of 19 October would have come sooner if the growing shyness of private lenders had not been obscured by the willingness of creditor central banks to keep on buying dollars, as required by the Louvre agreement earlier this year. Now, even the central banks are shy. Congress will have to do much better in fleshing out its agreement with the White House than in the negotiations leading to it if it is to persuade them to behave differently.

Meanwhile, the European Communities will yet again be seeking (and, almost certainly, failing) to assure their future by striking a deal, in Copenhagen, about the limits of what they spend on agriculture. Who can give much thought to the disappearance of assorted airliners, the impending famine in Ethiopia and the Irish question? Who, in these distracting circumstances, can dare assert his prior interest in the pursuit of natural knowledge? □

British government hedges bets on embryo research

- Long-awaited white paper pulls punches
- New licensing authority set up

London

THE British government has taken the first tentative steps towards legislating on assisted fertilization and embryology with the publication last week of a long-awaited white paper* (policy document) on the issues dealt with in the report of a committee chaired by Dame Mary Warnock, published three years ago (*Nature* 310, 266; 1984). The government has failed, however, to take a stand on embryo research — the white paper contains two alternative proposals, one banning research, the other permitting limited research on embryos up to 14 days old. Parliament will be given a free vote.

The white paper broadly follows the recommendations of the Warnock committee, with the central proposal being to establish a statutory licensing authority to regulate infertility treatment and, if it is permitted, embryo research. It does not follow Warnock's recommendation to legislate against non-commercial surrogacy. **Licensing.** A Statutory Licensing Authority (SLA) would license research or treatment involving human embryos created *in vitro* or procured from the uterus; treatment involving the use of donated gametes (for example, artificial insemination by donor) or donated embryos; the storage "in an arrested state of development" (currently by freezing) of human gametes or embryos for use later; and the use of diagnostic tests involving penetration by human sperm of an animal ovum. Licences would be renewable and granted for up to five years. It would be a criminal offence to undertake without a licence any of these activities.

The SLA's chairman and members (between 15 and 20) would be appointed, for a fixed, renewable term, by the Secretary of State for Social Services. The chairman and at least half the members would be lay people, and at least one-third doctors or scientists with relevant experience. The SLA would be funded partly by the state, but mostly by licence fees.

Embryo research. The draft clause prohibiting research reads: "It will be a criminal

offence to carry out any procedures on a human embryo other than those aimed at preparing the embryo for transfer to the uterus of a woman; or those carried out to ascertain the suitability of that embryo for the intended transfer". The clause permitting research begins: "Except as part of a project specifically licensed by the SLA, it will be a criminal offence . . ." In either case, storing embryos is allowed.

If research is permitted, the government proposes to accept the Warnock recommendation that a licence would forbid the use of embryos beyond 14 days after mixing the sperm and egg or after the appearance of the primitive streak, whichever is the earlier. Research licences would be linked to a specific project and would expire after a maximum of two years. For the SLA to grant a licence, it would need to be satisfied that the research was aimed at advancing diagnostic or therapeutic techniques of fertility control, was scientifically valid and that no reasonable alternative was available. The consent of the donors of the embryo or the gametes would be required.

The government's sensitivity to public opinion is illustrated by its rejection of the recommendation that the SLA should be responsible for issuing guidance on research unlikely to be considered ethically acceptable in any circumstances, such as cloning or genetic engineering. The white paper proposes that "the legislation should prohibit all such activities, but with the power for Parliament itself . . . to make exceptions . . .".

Control and storage. A licence from the SLA would be required to operate storage facilities for gametes or embryos whose control would be in accordance with the donor's wishes. The maximum storage time for embryos would be five years "or such lesser time as the SLA determines" and for gametes ten years. Transfer of gametes or embryos between facilities would require both parties to hold a licence, unless authorized by the SLA. The sale and purchase of human gametes and embryos would be controlled to avoid

Voluntary authority sighs with relief

London

THE body that at present monitors programmes of *in vitro* fertilization (IVF) and embryo research, the Voluntary Licensing Authority (VLA), has welcomed the white paper with relief. The VLA was set up in 1985 as a temporary measure. Its lack of statutory powers and the rapid evolution of techniques in the field has put its part-time membership and staff under increasing strain.

The VLA is presently considering requests from several IVF clinics to relax its guideline that relatives of infertile patients should not be allowed to donate gametes, on the principle that the anonymity of the donor is in the interests of the child.

Public interest in IVF has been heightened during the past few weeks, centring on the activities of the private Humana Hospital Wellington in London, which has Britain's largest IVF unit. The VLA withdrew its licence when the hospital refused to forbid the transfer of more than four eggs at a time. Since then the hospital has agreed to abide by the rules. The hospital also sacked from its ethics committee for breach of confidentiality a doctor who had supplied the VLA with details of the hospital's practices. Police are also investigating allegations that having transferred more than four eggs the unit selectively reduced the number of fetuses, possibly an offence of procuring an unlawful miscarriage.

Discussion of the ethical issues of IVF and embryo research will doubtless become entwined with the current public debate on abortion. A Liberal Member of Parliament, Mr David Alton, is seeking to change the time limit for abortion from the present 28 weeks to 18. Simon Hadlington

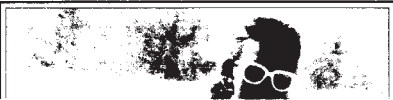
the risk of commercial exploitation.

Surrogacy. Most of the Warnock committee wanted legislation against private surrogacy arrangements. In 1985 the government made it an offence for a third party to take part in negotiations for a commercial surrogacy arrangement. The white paper says it would be inappropriate and not "necessarily in the child's best interests to make private, non-commercial surrogacy arrangements an offence."

All adults over the age of 18 would have the legal right to find out whether they were born following gamete or embryo donation, and the right of access to "certain non-identifying information". Where a child results from donated gametes or embryos, the carrying mother would legally be the child's mother. The donors would have no parental rights or duties.

Simon Hadlington

CLAMP ON FRANKENSTEIN SCIENTISTS



SCIENTISTS are to be banned by law from creating superbeings in the laboratory.

The popular press in Britain promotes a less than flattering view of scientists; Today's response to last week's white paper.

*Human Fertilisation and Embryology: A Framework for Legislation (Cmnd 259, HMSO, £3.30).

Proposal to put Prestel viewdata system into every British home

London

BRITISH Telecom has decided to embark on an ambitious scheme to put its pioneering viewdata system Prestel into every telephone subscriber's household. At present, only 24,000 British households have Prestel terminals, out of 18 million domestic telephone subscribers. It is understood that British Telecom will in the first instance set up a trial, with subscribers in one London telephone district and one district outside London being supplied with Prestel terminals for a fixed time. The town of Milton Keynes, a traditional test-bed for British Telecom's innovations and with the country's youngest, least technology-shy population, is being tipped as a possible trial district, although a city such as Nottingham would be more representative.

British Telecom has not yet publicly announced its plans, and the size of the trial sample is not known. It is understood that tenders have been invited for the supply of 5 million Prestel terminals, and Philips Business Systems is one of the companies invited to tender. Officials want the trial to start by September.

Launched in 1979, Prestel has never lived up to its early promise, and British Telecom's hope that the Prestel terminal would become as ubiquitous as the television set or refrigerator failed to become reality. Prestel was originally conceived as a means of increasing domestic telephone traffic, traditionally subsidized by the business community, but soon became established as a trade medium when the ferry company Sealink installed 600 terminals free of charge in travel agencies

to supply continuously updated information on ferry availability. Indeed, by 1981, British Telecom had changed its marketing strategy to aim at the business world rather than householders. Now, of Prestel's 78,000 terminals, 69 per cent are in the workplace and 31 per cent in the home.

The lack of interest of domestic users is generally blamed on poor marketing and the cost of hardware. At present, the least expensive means of using Prestel is to buy an adaptor priced at around £100 for an existing television set. Although Prestel's office says the system makes a profit (figures are not made public), of the numerous services available, it is the financial and travel companies that supply the bulk of the business.

Despite the relatively modest number of Prestel subscribers, the service is growing. The number of terminals has increased more than threefold since 1982-83 and in 1981 there were only 1,300 domestic users. If British Telecom does decide to provide householders with free Prestel sets, the capital outlay will be considerable. But one immediate benefit would be to relieve the much-lambasted directory service (the Prestel system would provide a national telephone directory), which is now free, although British Telecom hints that this might not always be so.

The overall standard of service from British Telecom has come in for a good deal of criticism recently. Its Prestel scheme could do much for public relations, and with a half-year pre-tax profit to September this year of £1,120 million, it should be able to take the financial strain.

Simon Hadlington

Racing for the cosmic flash

Sydney

ASTRONOMERS in Australia and New Zealand are rushing to complete gamma-ray telescopes in time to catch the expected arrival of gamma rays from the February supernova in the Large Magellanic Cloud.

Two new instruments are under construction; one in the Australian desert at Woomera and the other 1,650 m up a mountain near Blenheim on New Zealand's South Island.

Visually, the supernova has long passed its peak, but no gamma rays have yet been seen because they cannot penetrate the shell of hot gas and dust that surrounds the supernova's innermost regions. But that veil is growing thinner.

Gamma-ray telescopes observe indirectly. When gamma rays strike the atoms in the upper atmosphere, they scatter their energy in a cascade of secondary particles; these in turn give off flashes of light called Cerenkov radiation which can be picked up by the telescope.

The Woomera instrument, being built by the University of Adelaide's department of physics at a cost of A\$220,000, should be able to locate a light source to within a half degree on the sky. The telescope consists of 57 separate 80-cm diameter mirrors mounted on a steerable frame in three groups of 19. In each group, the light collected by the mirrors is brought to a common focus where it is picked up by a photo-electric detector.

The detectors can resolve time differences as small as five thousand millionths of a second. Any event not detected simultaneously at the three focal points is rejected as noise.

The Woomera site was chosen for its dark skies, the prime criterion for gamma-ray observation. The telescope had been planned long before the supernova and was not due for completion until the middle of next year. But the project leader, Professor John Prescott, now plans to have it operating by 1 January.

The New Zealand telescope owes its genesis entirely to the supernova. Universities in Japan, Australia and New Zealand are working together, with Professor Y. Murakami of the University of Tokyo as the project leader. According to Dr John Humble of the University of Tasmania, the telescope is largely being put together using spare parts gathered from around Japan. New Zealand is providing the site and Australia largely 'moral support'. The site was chosen in May, construction began in September and the first test observation had been completed by 20 November.

Charles Morgan

Seismologists unshaken by earthquake

Washington

Two more earthquakes hit southern California last week, both measuring a little more than six on the Richter scale. But unlike the earthquake of similar magnitude that rattled Los Angeles on 1 October, no deaths were caused and few puzzles were set for seismologists.

Robert Wesson, a geophysicist specializing in earthquake prediction at the United States Geological Survey (USGS), said that while the 1 October earthquake "made us feel humble, this one made us feel that we are beginning to understand what is going on". The pair of earthquakes occurred on the San Andreas fault in the sparsely populated Imperial Valley close to the Mexican border.

The area's location at the end of the San Andreas fault system puts it at risk from earthquakes. Over the past decade, there

has been a pattern in the place and time of Californian earthquakes that, while not fully understood, suggests that stress, generated in the slow movement of the Pacific plate, is transferred up and down the faults in a predictable manner. Stress is particularly likely to concentrate at the end of faults which are slowly creeping. Superstition Hills had been marked down by USGS researchers as a possible spot for an earthquake during 1986-96. Earthquakes occurred nearby in 1979 and 1980.

The citizens of Los Angeles are not likely to gain much comfort from the knowledge that long-term earthquake prediction is gaining some respectability. Similar studies by the USGS suggest that there is a 50 per cent chance of a magnitude 8 earthquake, likely to cause 3,000-13,000 deaths, on the San Andreas fault in the next 30 years.

Alun Anderson

France prunes research spending but supports big science

Paris

AFTER two years of internal turmoil and a budget increase below the rate of inflation, France's largest science research body, the CNRS (Centre National de la Recherche Scientifique), has taken the step of giving priority to big science. As is usual nowadays, however, the money is to be found by pruning new research jobs and by seeking industrial sponsors.

"It's not an easy budget", said Serge Feneuille, director-general of CNRS, as he outlined the priorities for 1988 on 23 November. Spurred on by the government, the CNRS has decided to focus support on its best laboratories, to ensure that France continues to compete internationally. "This will mean some painful decisions have to be made", said Feneuille, "but no discipline will be sacrificed".

The European Synchrotron Radiation Facility (ESRF), at Grenoble, is to receive FF45.5 million (£4.6 million) in the coming year, a decision that must be linked to France's unwillingness to participate in Britain's spallation neutron source, ISIS, at the Rutherford Appleton Laboratory (see *Nature* 329, 381; 1987). The Vivitron at Strasbourg will also receive FF19.4 million in 1988.

A number of new initiatives were announced by Feneuille, including provision of FF6 million for the Very Large Telescope (VLT), planned for a site in Chile. Commitment to VLT from other European partners has not yet been found but, according to Feneuille, "we are going ahead as if it were a certainty". The money will go to interferometry and mirror-polishing laboratories, where France has a solid reputation. Similarly, high-temperature superconductor research will get an extra FF8 million, and information technology an extra FF5 million.

Two new laboratories will be created in life sciences, one in plant biology, at Gif-sur-Yvette, near Paris, and the other in pharmacological biotechnology.

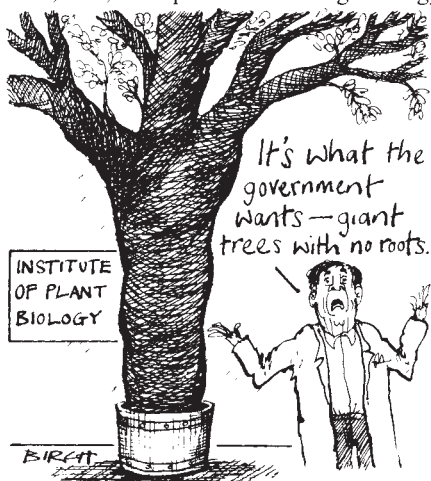
This veneer of progressive dynamism conceals a heavy reliance on contributions from industry, as could be anticipated from research minister Jacques Valade's budget speech in the Senate on 21 November. Valade called the collaboration between industry and public "the priority of priorities" and sees it as a way not only to co-finance, with CNRS, around 50 research laboratories over the next five years, but also to make up for a stagnation in researcher training and recruitment (see *Nature* 329, 380; 1987).

An extra 40 doctoral training scholarships will be available to engineers in the fields of chemistry and physics, if industrial co-sponsors can be found to match

the FF10 million the CNRS has earmarked for this purpose. About 200 of these scholarships are awarded each year. Similarly, 20 two-year postdoctoral fellowships will be available on the same partnership basis with the private sector. The reservation of these opportunities for research areas with an obvious technology bias is related to the government's new scheme of financial incentives (a year's salary) for trained scientists leaving the CNRS for industry.

The defence budget, given a 10 per cent increase for 1988, will also be tapped. The French military research establishment, DRET, has a 25 kW high-energy laser which will form the basis of a joint laboratory with researchers from CNRS studying metallurgical applications.

The linear accelerator now at the University of Paris (Orsay) will move to a new, 10,000-square-metre high-energy



physics laboratory at the Cité Descartes, a science park at Marne-la-Vallée, the new town under construction just outside Paris. "It will be good to have two high-energy physics groups in Paris", said Feneuille, an atomic and molecular physicist, "a bit of competition is constructive".

A comparison of French expansiveness with increasing British reticence towards European big science areas will be inevitable. In his Senate speech, Jacques Valade admitted that "it has become fashionable to caricature France's position", but this could also be said of Britain. In reply to a question about Britain's role in ESRF, Feneuille said "the British scientific community must be present, since they have some of the best researchers in the world, especially in molecular biology and protein crystallography. But it would be difficult to accept that they are present at cut-price, on the same level as a small country".

Peter Coles

Britain drags feet

London

ENVIRONMENT ministers from eight European states have agreed to adopt a series of measures to reduce pollution in the North Sea. Meeting in London last week, ministers from Belgium, Denmark, France, West Germany, the Netherlands, Norway, Sweden and Britain decided to end incineration at sea by 1994 and halve by 1995 the volume of toxic chemicals discharged.

Britain would not agree to cease dumping sewage sludge, but gave a commitment at the meeting that the levels of contaminants would not be further increased. Dumping of industrial wastes that cannot be shown to be harmless will be phased out by the end of 1989. It was agreed to ban the dumping of garbage from ships, to increase aerial surveillance of the sea and to set up a scientific task force to monitor and assess pollution.

S.L.H.

Go-ahead for BRAIN

Brussels

WITH its penchant for memorable acronyms, the European Commission has approved a budget of £690,000 for six neurocomputer projects under the title of BRAIN (Basic Research in Adaptive Intelligence and Neurocomputers). The projects, regarded as the European Community's response to Japan's Human Frontier Science Program, will involve around 100 researchers in 28 laboratories in West Germany, Britain, France and the Netherlands.

B.C.

Indifference bemoaned

London

More in sorrow than in anger, Sir George Porter, president of the Royal Society of London, this week wrung his hands over the indifference of the British to "the pursuit of knowledge for its own sake". He was delivering this year's anniversary address to the Royal Society.

Porter said that, while there are signs that the United Kingdom is prospering in some respects, it seems that "the pursuit of natural knowledge is to be allowed to diminish". Pleading that the argument about the state of British science should not degenerate into demands that "the government should hand out more money", he said the "deeper and more serious" trouble is that "neither governments nor the country" seem to have much interest in research.

Porter went on to describe how he had been told by ministers and civil servants that there is "too much science", that Britain "can leave it to others" and that the "importance of Nobel prizes went out with Harold Wilson". He said that the "only argument for improving natural knowledge which carries any weight today is that it creates wealth, and even this is refuted in some quarters".

□

Summit meeting may give French videotext international appeal

Paris

THE hugely successful, but as yet largely unexportable, French public videotext system, Teletel, turned a historic corner at the beginning of November when West Germany finally agreed to allow connection of Teletel's 'Minitel' terminals to the public telephone network. The agreement, signed by telecommunications ministers from France and West Germany at a summit meeting in Karlsruhe, followed a long dispute over standards.

West Germany's principal objection to Minitel — its internal modem — was gradually eroded during negotiations. German modem manufacturers (such as Siemens) wanted Minitel to use external (German) modems.

West Germany is now the first European country to connect directly into the French videotext network, effectively doubling the potential number of users. As part of the deal, French telecom (DGT) will team up with its German equivalent, Deutsche Bundespost, in the design of a new telephone workstation and the construction of a fibre-optic link between Mulhouse and Karlsruhe.

Since the first field trials between 1980 and 1983, Teletel has attracted almost 3 million subscribers, whereas Britain's Prestel — which pioneered videotext and may yet prosper, see page 410 — still has only 78,000 and West Germany's Bildschirmtext (BTX), 60,000. The French DGT learned from Prestel's slow take-off that the public were reluctant to invest in the hardware needed to connect into the network, particularly in the beginning, when few services were on offer. Similarly, with few subscribers, it was hard to persuade advertisers and suppliers of valued-added services of the potential.

The DGT therefore decided to create a large pool of Minitel users at the outset, by distributing free terminals (comprising internal modem, screen and keyboard), with free access to an electronic directory service (EDS), instead of the telephone book. EDS, the only Teletel service offered by DGT, is still the most called, registering 24 million calls and 900,000 connection hours in June this year.

The success of Teletel is due also to a kiosk system of charging for services, whereby suppliers take part of a higher telephone charge, instead of demanding a subscription fee, which often discourages casual users. A flourishing range of online bulletin-boards has grown up precisely because callers remain anonymous.

The modular conception of the underlying communications network also makes it efficient and easily expandable. While user terminals are connected to the ordi-

nary public-switched telephone network, host computers supplying Teletel services are connected to the Transpac packet-switched data network. The link between the two is provided by access points around the country. As Transpac provides gateways to foreign telecom networks around the world, Minitel may, in theory, be exploited abroad.

Here, however, comes the rub. Teletel databases are of interest to foreign users, while the large French user network is attractive for foreign service suppliers. The basic Minitel terminals, manufactured by Telic Alcatel, Matra and Radiotechnique, are, however, too rudimentary to warrant substantial export orders. In five years, only 100,000 terminals have been sold abroad — notably to Spain (30,000), Switzerland (6,000), Canada (30,000), the United States, Belgium and Brazil. These terminals are usually used as

part of an internal network, not to connect with French Teletel.

Differences in telecom standards between France and other countries have also limited attempts to export Minitel technology. The standard Telic Alcatel/Radiotechnique Minitel uses an outdated 'V23' modem that does not conform with rapid data-transmission standards in the United States and some other countries. So multistandard 'de-luxe' Minitel, including a joint Teletel/Prestel version, have been developed by French manufacturers, with approval in 14 countries.

However, the growing number of Minitel emulation cards for personal computers may nip in the bud the expected extra business for Minitel's French makers while cutting down DGT's profits from Teletel calls. These cards make use of hardware already purchased by the user, while adding a modem. Clever software uses the computer's peripherals to shortcut the time-wasting advertising pages and inefficient Minitel screen, which usually prolong EDS calls beyond the first 3 minutes for which they are free. **Peter Coles**

NSF to spend \$14 million on MERIT supercomputer network

Washington

MERIT, a consortium of Michigan universities, will soon be taking over the operation of NSFNET, a high-speed, nationwide backbone network that links regional computer networks and supercomputer centres. The National Science Foundation (NSF) announced last week that it is giving MERIT a grant of \$14 million for five years to run the network. But also important to MERIT's plans to expand and upgrade NSFNET are commitments of financial support and support in kind from IBM, MCI Communications Corporation and the state of Michigan.

Computer networks have proliferated in the United States, creating protocol nightmares for internet communications. NSFNET is an attempt to simplify matters.

Sending data through a network is accomplished using packet switching; data files are broken into individual 'packets', routed independently through the network in the most efficient manner possible and then reassembled at their destination. The speed of the network is determined by how fast the transmission lines connecting the nodes can transmit data, and how clever the algorithms are for routing packets efficiently through the network.

MERIT's plans for NSFNET call for transmission lines that can carry data at 1.5 megabits per second to be in use by June 1988. MCI will provide long-distance lines and will help underwrite the cost of local telephone company connections.

MCI's contribution is reckoned to be

worth approximately \$7 million. For its part, IBM will provide switching hardware, as well as software for network management and informational services. Depending on how it is calculated, IBM's input to NSFNET is valued at between \$10 and \$30 million. The state of Michigan is contributing \$5 million to the project.

Stephen Wolff, director of NSF's division of networking and communication research and infrastructure, sees NSFNET as an opportunity for industry to get a head start on what he expects will be a proliferation of high-speed data networks in the future. He also hopes NSFNET will help NSF to take the lead in making sense of the current jumble of networks used by federal agencies.

NSFNET will be operational next summer using 1.5-megabit lines, but MERIT plans to develop the switching capability to handle lines that can transmit data 30 times faster. Such high-speed networking capabilities open new possibilities in real-time communications. Higher-speed networking also will mean greater remote access to national supercomputer centres. NSF now supports five such centres: Cornell University, Princeton University, the University of Illinois at Champaign-Urbana, the University of Pittsburgh and the University of California at San Diego. Each will be a node in the NSFNET backbone together with the supercomputer centre at the National Center for Atmospheric Research in Boulder, Colorado. **Joseph Palca**

Networks in the United States

Washington

THE US government will need to double its current spending on high-performance computing if it is to retain strength and leadership in that area for the forthcoming decades. That is the conclusion of a report requested by Congress and prepared by an interagency committee of the government. The government now spends some \$500 million per year supporting research and development in computer hardware, software and networking capabilities.

Although market forces will push commercial development of computing capacity, resources that can meet the special computing needs of fundamental scientific and engineering research will require a "shared vision" among government, industry and higher education for the proper allocation of resources.

The new report was written by three subcommittees of the Federal Coordinating Council on Science, Engineering and Technology (FCCSET) committee on computer research and applications, whose membership included federal agencies with interest in computing activities. Paul Huray of the University of Tennessee, who chaired the FCCSET committee, says the report represents a first step in reaching that shared vision.

The FCCSET report was originally intended to address only computer networking, but Huray says the committee felt that overemphasis on networking might obscure needs in other areas of computing. In hardware, the report urges a long-range strategy for federal support for basic research, and improved mechanisms for transferring government-sponsored technology into the private sector.

At a time when the White House has been looking for ways to trim the federal deficit, it is notable that the FCCSET report, reviewed at the highest levels of the White House, calls for increased federal spending. The report recommends an immediate doubling of the \$50 million spent annually on networking, with an incremental rise in federal spending on other areas of advanced computing leading to a doubling of current spending in five years.

The 29-page FCCSET report, originally a much larger document, deals only briefly with issues that are sure to be contentious. Who should pay for using the research network? What government agencies should be in charge of supporting basic research? How should government help industry to compete in global markets? Huray says the FCCSET report is intended to stir up those questions, rather than suggest formal answers.

Joseph Palca

Doubts about customers for planned Japanese network system

Tokyo

NIPPON Telegraph and Telephone (NTT), Japan's giant domestic telecommunications company, is pressing on with plans to introduce a commercial digital Information Network System (INS) linking Tokyo, Osaka and Nagoya by early next year. But the results of a trial in Tokyo suggests that there will not be a flood of INS subscribers.

In September 1984, NTT began an ambitious experiment to test a model INS in the Mitaka suburb of Tokyo. About 2,000 subscribers (both household and business users) participated in the two-and-a-half year project which cost ¥20,000 million (\$150 million).

The idea of INS is to link telephone, telex, facsimile, teletext, computers and video transmission to one communications network. Mitaka subscribers were allowed to test out all the latest electronic gadgetry, including videophones, sketch-phones, high-speed digital facsimiles and the CAPTAIN videotext system.

But customer reaction to the new information age is mixed. Although high marks were awarded by companies to videophones, video conferences and high-speed facsimile machines, housewives complained that their high-speed faxes were inundated with junk mail from local supermarkets (see below). CAPTAIN terminals were too difficult for some of the general public to operate and users had to wade their way through pages of uninteresting information on the screen — a common criticism which will only be overcome when much more sophisticated software is developed. And Nobuo Inoue, executive director of NTT's research and development headquarters, admits that he has "no idea" how to deal with the

problem of "information overload".

NTT is nevertheless pressing ahead. A larger field experiment covering Tokyo, Osaka, Nagoya and the academic city of Tsukuba has already begun with the participation of 20 private companies. And NTT plans to convert the system to a commercial service in March 1988.

According to Inoue, one of the most popular experiments in Mitaka involved the link-up of satellite offices in the suburb with main offices in the downtown area of Tokyo. A branch office of NEC, the electronics giant, and a printing company with a printing works in the suburb found that by using a videolink, employees could avoid shuttling back and forth between the main and satellite office. If such a system could help ease the congestion of Tokyo's overcrowded commuter trains it would certainly be welcome.

Broadband videolinks, however, are still far too expensive to contemplate. But NTT has developed a second system that can transmit real-time pictures, provided the conference participants do not move around too much. And Inoue thinks this is commercially viable.

So how much will INS cost the subscriber? NTT is negotiating with the Ministry of Posts and Telecommunications. NTT officials admit that nobody has a clear idea of what would be an appropriate tariff system — a "political decision" is required, they say. But NTT expects that, at least to begin with, INS will be four or five times as expensive as the present telephone network.

Judging from the reaction of the general public in Mitaka, it seems safe to conclude that for the next several years INS use will be confined to corporation that can afford to pay the bill.

David Swinbanks

Direct mail ordeal by facsimile machine

Tokyo

JUNK mail by telefax? The information age has arrived in Japan with a vengeance. NTT's trial of an Information Network System in a suburb of Tokyo had one unexpected result: housewives were deluged with unsolicited promotional material received on high-speed facsimile machines installed in their homes. This new form of direct mailing threatens to become a major nuisance for facsimile owners — and there are now 1.5 million machines in Japan.

The very high postal charges in Japan make this new form of direct mail particularly attractive, and direct mailers are not the only ones to see the potential of facsimile machines. Some pranksters dial up facsimile machines at night and feed a

continuous loop of paper into their machine with indecent or abusive words written on it. The next morning the unfortunate recipient finds his facsimile emptied of paper and the floor strewn with messages. As a continuous loop is used there is no record of the sender.

How can facsimile owners be protected? Dr Nobuo Inoue, executive manager of NTT research and development headquarters, says that facsimiles could be programmed to receive only from specified numbers. A more practical solution may be to design facsimile machines that can initially store messages electronically and display them on a screen. The recipient could then print out only selected material.

David Swinbanks

Hundreds of Californian species sliding towards extinction

San Francisco

Two hundred and twenty animal species and 600 plant species in California are "sliding toward extinction", according to the first comprehensive assessment of the state's species and natural communities, recently released by the California Nature Conservancy. The report, requested by the state legislature for use in evaluating and designing conservation legislation, found that shrinking habitats are the main threat to endangered species, which include 10 per cent of the state's native mammals, 17 per cent of native reptiles and amphibians and 27 per cent of freshwater fish. The report proposes new legislation, including tax incentives and habitat protection, to halt the decline.

California is in a unique position, according to Steve McCormick, director of the California Nature Conservancy. The state not only has more species diversity than any other area of the country because of its size and range of topography and climates, but culturally, economically and politically it is in a position to "take the wise path" towards saving its wildlife.

The report criticizes California laws that do not protect endangered species from being killed on private land, and

state agencies that, without adequate funding, are not able fully to evaluate species or to review development projects.

Among its recommendations, the report proposes acceleration of state programmes to identify and acquire endangered habitats for protection, extension of the protection of endangered species onto private land and new laws to protect endangered species and their habitats.

The report also suggests tax-incentives to encourage landowners to return their land to the wild state. McCormick says these would work well for wetlands, important habitats for elk, deer and wintering water birds, which are easily restored by dismantling dams and levee systems.

California wetlands have shrunk to 4 per cent of their original extent, due to conversion to farmland, diversion of water and pollution by agricultural runoff. In 1983, the California legislature vowed to restore wetlands to 150 per cent of their present area by the year 2000.

No attempt has been made to estimate the cost to the state of implementing the report's recommendations, and there is no indication whether any of the recommendations will be pursued by the legislature.

Marcia Barinaga

Soviets ponder AIDS vaccine

London

THE Soviet Union has recently stepped up research to find an anti-AIDS vaccine or prophylactic, according to Academician V.I. Pokrovskii, president of the Soviet Academy of Medical Sciences. Speaking last week at a televised news conference, he warned that the HIV-1 virus is now "at large" in the Soviet population, and has now reached people who have had no direct contact with foreigners.

Pokrovskii spoke of "around 32-33" Soviet citizens infected with HIV (human immunodeficiency virus), including one case of fully developed AIDS, plus an unspecified number of foreign students. "The first investigations" for HIV-2 have now started, Pokrovskii said.

Glasnost came late to AIDS, perhaps as a result of pre-Gorbachev propaganda which labelled it a disease of capitalism and/or the result of a bacteriological warfare experiment in the United States that went wrong. Testing Soviet citizens has considerable logistic difficulties as prostitution and homosexuality are illegal.

Voluntary anonymous testing facilities for HIV have been available for some time, but a recent decree on mandatory testing of suspected carriers and the heavy legal penalties it imposes on carriers who knowingly place others at risk make it difficult for many people in the high-risk category to trust fully in the promised confidentiality. Moreover, if the Soviet Union continues its policy of compulsory testing, it could encounter difficulties internationally. As Pokrovskii stressed, international cooperation is all-important in AIDS research. But the Hungarians who had also passed a new law on the mandatory testing of high-risk groups, now say that they may have to abandon it in the light of the agreement rejecting compulsory screening adopted by a recent conference in Paris.

In the meantime, research efforts continue, and receive considerable publicity. The popular science journal *Nauka i Zhizn* recently described the development of 'nuclear filters' at the Dubna Joint Nuclear Research Institute. These filters, it was explained, are produced by directing a beam of charged particles at a polymer film, and can be tailored to trap the HIV virus, which would interact with the material of traditional member filters. Two weeks ago, the official news agency TASS reported that a joint research project had been agreed between Finnish scientists who have produced individual parts of an HIV-protein using genetic engineering and a Soviet team who had synthesized immunity stimulants to produce a vaccine against HIV.

Vera Rich

NSF's technology fans

Washington

PLANS by the National Science Foundation (NSF) to establish centres for the exploitation of science and technology (*Nature* 328, 5; 1987) have drawn an enthusiastic response. Nearly 900 'letters of intent' have been sent to the NSF, mostly by individual universities but in some cases by consortia of universities, industries, national laboratories and state governments. Alan Leshner, in charge of NSF's programme for the science and technology centres, says that the variety of proposals and proposers betokens an "interesting and exciting" approach to the problems the research community faces in times of reduced funds and worries about foreign superiority.

To support all the proposed centres, NSF would need \$10,000 million over five years. So far, NSF's request for \$30 million in the first year lies, with all other federal money, in budgetary limbo while Congress and the president try to agree on reducing the deficit reduction. But this programme is likely to get sympathetic attention. In the meantime, Leshner says that NSF will go ahead with evaluation and review, reducing the 875 applications to a shortlist of perhaps 40 in anticipation of financial support.

David Lindley

Britain's space plea

London

MR Jack Lemming, the new director of the British National Space Centre, has supported the British government's refusal to join the expanded European space programme. His predecessor, Roy Gibson, resigned earlier this year.

Lemming said last week that, although Britain rejected outright participation in the space aircraft Hermes, it could still participate in a range of projects on which "we have not got a clear policy". He called on industry to offer support to the space industry and to boost Britain's stake in the European programme.

Reaction from industry may be mixed. Contracts worth at least £2 million have already been lost by British companies following the decision and they have also been told that they cannot bid for future work on Hermes, worth hundreds of millions of pounds. Under European Space Agency rules an industry is not eligible to work on projects which the home government refuses to support.

Lemming, a career civil servant, is due to retire at the end of this year. He nevertheless hopes to complete the campaign to involve industry and the scientific community before then.

Sarah Hargreaves

Expanding Research Triangle draws new land, new participants

North Carolina

GLAXO, the British pharmaceutical company, has apparently discovered something North Carolina officials hope US companies will start discovering in greater numbers — the triangle of land that has the University of North Carolina, North Carolina State University and Duke University at its corners contains one of the highest concentrations of PhD scientists in the country.

This has made the region fertile ground for university, industry and government collaborations to turn academic brain power into high-technology products.

The proliferation of 'research parks' is a new phenomenon, but Research Triangle Park in North Carolina is now almost 30 years old. It has been through some ups and downs as political currents have shifted (see below) but now contains 54 companies and government agencies, including Burroughs Wellcome, IBM and Northern Telecom, all within a 15-minute journey of the three universities. And although the park is the largest in the world it is still growing — a new southern zone will bring the park's area to 6,700 acres.

New companies are still coming to the park and new kinds of associations are being formed among the three local universities, industry and government. At the University of North Carolina (UNC) at Chapel Hill, the pharmaceutical company Glaxo has entered into a unique arrangement that will allow university scientists and Glaxo researchers to share a \$3-million research facility.

The institute is the brainchild of Glaxo's research director Pedro Cuatrecasas. After moving from Burroughs Wellcome,

he wanted to build up a solid research team quickly, without waiting for Glaxo's massive new US headquarters to be completed in the park. Instead of opting for a temporary laboratory, Cuatrecasas approached UNC-Chapel Hill with an offer to build a research laboratory on campus that would revert to the university at the end of five years, or whenever



Glaxo's headquarters was finished.

Research will centre on cancer and viral diseases, AIDS (acquired immune deficiency syndrome) and molecular genetics — "very similar to the kinds of research already going on at UNC...but more focused on the use of new chemicals and organic chemicals", says Cuatrecasas. There will be no restrictions on the publication of research performed in the facility, and Cuatrecasas says that "it's possible we will make mistakes" in not retaining property rights over substances developed

there. But no drug development will occur in the building, and he says the risks of losing exclusive rights over new substances are more than compensated for by the ability to conduct research in an environment that promotes the free exchange of ideas and mutual trust. Glaxo scientists will hold teaching posts at the university.

Cuatrecasas believes that the exposure of university scientists and students to people who are actively pursuing therapies will give them a better perspective on their own research, and possibly make them aware of alternative career paths.

Over at North Carolina State University, the university is celebrating its hundredth birthday by building a new research campus where academic and industrial scientists can work together. The first building is nearly complete on a 780-acre site granted by the state. Research will be concentrated in traditional strengths: textiles, engineering, and agricultural and forestry biotechnology. Included will be a conference centre and hotel, residences for visiting scientists, and 85,000–1,150,000 square feet of office space. Part of the space will provide 'incubators' for new companies intending to exploit technology developed on campus.

Development funds will come from the state and through bond issues, but Philip Carter, associate vice-chancellor for research, says the university will also take a "let's make a deal" approach in accepting offers from companies wishing to finance a research facility. Many of the projects undertaken on the campus will bridge the gap between basic and applied research, and cut across traditional disciplinary boundaries. The nearly completed 30,000-square-foot engineering building will house research on precision engineering and semiconductors, with 5,000 square feet of space reserved for industrial partners. A similar life-sciences building will be constructed next, with an emphasis on biotechnology.

Much of the 'incubator' space will be partially funded by the North Carolina Biotechnology Center, set up to boost biotechnology in the state. A total of \$6.5 million will be given this year to academic researchers and to help new companies get on their feet. Among the companies set up in the centre's six-year history is Embrex, Inc., a poultry biotechnology company started in 1985 using technology licensed from the US Department of Agriculture. The company has developed an injector that should make it possible to replace the costly business of vaccinating hatchlings against the poultry disease coccidiosis with automatic vaccination of embryos in the egg.

Carol Ezzell

Thirty years of Research Triangle Park

North Carolina

THE decision to create the Research Triangle Park in 1957 was prompted by a state-wide study to find ways to broaden North Carolina's tax base and to staunch the 'brain-drain' of promising young scientists graduating from the local universities and leaving the state. The state sought to lure chemical, pharmaceutical and electronics companies to the area, and local faculty called on corporate executives to sing North Carolina's praises.

In the autumn of 1958, \$1.5 million was raised in 60 days, \$1 million towards the development of the property and the rest towards the foundation of Research Triangle Institute, a private research institute set up to do contract research.

The first company to build in the park

was Chemstrand, later to become Monsanto, in 1959. Smaller companies began to trickle in, but North Carolina's racial problems in the early 1960s kept companies from investing in the area.

The next major coup for the park came when former state governor Luther Hodges, then US Secretary of Commerce, persuaded the president of IBM to take a look at the area. IBM built a research facility in the park in 1965, the same year the US National Institute of Environmental Health Sciences was located there, and Research Triangle Park was on its way.

In recent years, the US headquarters of Burroughs Wellcome has been built in the park, and Glaxo is building an extensive facility there at its headquarters, expected to be completed in 1991.

Carol Ezzell

NIH budget: correction

The budget of the US National Institutes of Health (*Nature* 329, 573; 1987) is \$6,200 million.

□

The case for high-energy physics

SIR—Is the *impression* of being second rate of greater concern than the actuality? That is one of the conclusions that may be drawn from the article by John Maddox (*Nature* 329, 759; 1987) musing on the question whether Britain should withdraw from CERN (the European Organisation for Nuclear Research, see also p. 433).

The article advocates that Britain should pay its subscription yet refuse to provide high-energy physicists with the sums said to be needed "to make full use of the opportunity created by CERN membership". Such a course would not salvage our reputation as first-class scientists; we are assessed by results, not by whether or not we stick to treaties.

In such an atmosphere, few new physicists would risk their careers in our field, and so it would dry up. Once dead, it is hard to imagine it ever being reborn. Some endeavours simply cannot be run by market forces — among them, it seems, even markets themselves.

There is no point in debating whether money liberated from high-energy physics would indeed go to support the host of undeniably worthwhile projects in biology, chemistry and other fields of physics now reprehensibly being starved of funds. This possibility may have been fostered deliberately to divide the scientific community and thereby to make cuts within science seem self-inflicted.

The present assault on high-energy physics comes not so much from a rational assessment of society and civilization's long-term needs, but from a more general lack of comprehension, by this and previous governments, of the role of science in the modern world, which have picked the easiest targets — astrophysics, particle physics and even oceanography.

There is also little point in indignation at the strange assertions in the article, surely intended to convince the reader that high-energy physicists are profligate and financially irresponsible in these stringent times. In fact, CERN decided on LEP ten years ago, and is not "embarking" on its construction. And LEP was never intended to look for "string particles", offered as the theoretical star guiding us experimentalists. Its aims are far less esoteric and much more worthwhile. LEP is generally acknowledged to be the machine of choice for testing the Standard Model to its limits.

To claim that the benefits of LEP have never been proclaimed is untrue. Of course we know "what high-energy physics is for" — it is to study the structure and forces of the Universe at extremely small distances or equivalently high energy. What we don't know are the answers: if we did, but only then, high-energy physics would be redundant.

To dig beneath the "huff and puff", I can tell you why high-energy physics is a justifiable intellectual pursuit. It is one of the several areas of science whose goal is to try to extend humankind's knowledge of the Universe and of our role therein. By probing structure to ever-smaller distances, we are studying effects and phenomena that prevailed at the beginning of the Universe, and which have determined its development ever since.

By digging deeper, we discover that seemingly disparate phenomena are really manifestations of the same underlying truth, much as Maxwell's work embodied the unification of electricity and magnetism. The generally benign effects of electromagnetic technology in society have been a direct result.

The technological fruit of pure research is not, however, strictly relevant. High-energy physicists believe that a vigorous search for knowledge brings its own rewards, both for individuals and society. To renounce the inquisitive drive is to accept our present prejudices and misapprehensions. It is to renounce the ideals of creative curiosity and consumerism, it is less from arrogance and more from our perhaps naive lack of comprehension that people might not find it as beautiful a subject as we do. And then there is the fact that staying at the front requires all one's energies.

JOHN HASSARD

Blackett Laboratory,
Imperial College of Science and
Technology,
London SW7 2BZ, UK

John Maddox writes: The article complained of concluded that "the British government's best course is to stay a member of CERN, but to let high-energy physics grants take their chances with others at peer-review meetings". □

First World guilt

SIR—Any possible responsibility of the United States in the poverty of underdeveloped countries was denied by Jocelyn Tomkin (*Nature* 329, 10; 1987) with the naive argument that, should America disappear from the Earth, "World market prices of raw materials and crops would collapse, a major source of sustenance for third world countries would be withdrawn and, far from rising out of poverty, they would sink into deeper poverty". Exactly as the fact that babies need parents' help does not demonstrate that parents are not or cannot be violent against babies, the fact that the First World buys things from the Third does not demonstrate that it cannot be responsible for its poverty. Furthermore, Tomkin did not add any

reference to support the implicit statement that the 'First World' is of benefit to the Third.

It is true instead that developed countries receive from underdeveloped ones more than they give, so that world's money flows upstream from the poorer to the richer. This is made possible, in spite of the helps and loans that the First World gives to the Third, by the huge amounts of money our industries, companies and agroindustrial complexes earn from low-cost plants sited in developing countries^{1,2}; let me only quote the Union Carbide plant in Bhopal, which has become tragically known only by chance.

If the advantages the First World obtains from the Third were not so large it would be very difficult to explain why countries rich in natural resources (petroleum, metals, diamonds, agriculture) can be very poor.

It is impossible to discuss here the economic relationships between developed and underdeveloped countries, but there is little doubt that underdevelopment is a cultural reality more than an economic one. The end result of a cooperation in which the Third World gives its non-renewable resources and the First its renewable technology is that the United States and Europe consume one third of the total products.

Maybe the scientific community cannot do very much to modify these large-scale economic equilibria, but it is important at least to be aware that developed countries are partially responsible for the poverty of developing ones.

ANDREA BELLELLI

P. Trasimeno 2, 00198 Roma, Italy

1. George, S. *Comment meurt l'autre Moitié du Monde* (Laffont, Paris, 1978).
2. Gentilini, M. & Duflo, B. *Medicine Tropicale* (Flammarion, Paris, 1977).
3. Cross, N. *New Scientist* 114, 34 (1987).

Mistaken identity

SIR—Dreams really might come true if certain persons are in "competition to emulate President Theodore Roosevelt's fatuities on the similar [stock market crash] events of 1929" (*Nature* 330, 1; 1987): the poor chap had been dead for 10 years! President Reagan's other, more vital, counterpart in the crash of 1929 was President Herbert C. Hoover (no relation to Edgar J.), who was defeated in 1932 by Franklin D. Roosevelt who, in turn, went on to launch the New Deal and was handsomely re-elected for his efforts in 1936. Would that Senator Robert Dole, despite his lack of famous namesakes, have similar success to Roosevelt (the vital one) in his proposed presidential legislative programme for 1989!

PIERS R. ALLEN

JAI Press, 3 Henrietta Street,
London WC2E 8LU, UK

New European collaborations

The European Science Foundation has an enviable record in breaking European national chauvinism, but is now having to turn away good proposals. It needs a much bigger budget.

AMONG grant-making agencies, the European Science Foundation enjoys the distinction of spending the world's most weightily considered research budget. Last week's general assembly, which in the end approved a budget for 1988 amounting to FF 11 million, just over £1 million, brought to this inaccessible city the presidents of half a dozen European national academies and research councils together with personages almost as exalted representing other national research councils and international organizations.

For people who are responsible for annual budgets several hundreds of times larger (those of the West German *Max-Planck Gesellschaft* or the French Centre National de la Recherche Scientifique, for example), the case for spending two days travelling to and from Strasbourg cannot simply be the food. The truth is that the European Science Foundation (ESF) seems to be every participant's concept of what constitutes a good idea. The budget may be small, but it is not the future pattern of European research in microcosm?

But ESF is neither a research organization, with laboratories or research employees of its own, nor a foundation, with money of its own to spend. Instead, it is a membership organization whose members are national academies and research councils and whose dues constitute ESF's sole source of funds. A further complication is that the 49 member organizations come from a set of 18 European countries that seems not to coincide with any other — as well as the members of the European Communities, for example, there are three Scandinavian countries, Austria, Switzerland and Yugoslavia.

ESF has several ways of working. Some of what it does is hortatory, as when it produces documents such as that nearly ten years ago regretting the neglect of space science by the European Space Agency. Sometimes it functions as a midwife, as when it set out to make the case for a European synchrotron radiation source (and then found itself at the centre of a squabble about the siting of the machine, which is now being built at Grenoble). In these roles, which are supported from the general budget, the foundation is evidently a convenient stalking-horse for at least some of its members: it may be able to say things that would get individual members into trouble.

Last week's general assembly approved

an extension of the European Geotraverse project until 1990, partly because those who have collected data from the various sections of the traverse believe that time could usefully now be spent on the coordinated interpretation of the data but also because the European Commission has just made a grant of 1.2 million ECU (European Currency Units, about £800,000) to support a continuation of data-gathering in the Iberian peninsula.

Although ESF's direct spending on this project (from the funds contributed by members electing to take part) is relatively modest, covering the cost of a project officer based in Zurich and various workshops and planning meetings, its members in their role as national research councils have spent many times as much on research support for the groups that have carried out the work. Some of those who travel to Strasbourg say that their justification is the potentially high leverage of the foundation's small budget.

Additional activities will continue to be added. Last week, the assembly agreed to the first phase of a project, urged by Dr Peter Fricker, secretary-general of the Swiss National Science Foundation, to compile a coordinated palaeoclimatology of Europe based on physical, archaeological and historical records; a scheme for a coordinated study of the emergence in Europe of the modern nation state (dated at the thirteenth century); and a scheme labelled "environmental toxicology" which began life as a wish by member organizations for an objective basis on which their governments might frame environmental legislation, but which had some people last week sucking their teeth because of its imprecision.

But this year's new flavour is called "networks", and springs directly from the recognition that, if the enduring value of what the foundation does is to demonstrate to researchers the value of collaboration, the simplest way of maximizing this benefit may simply be to bring people together to discuss common problems.

Sir Arnold Burgen, until recently foreign secretary of the Royal Society, has been the chairman of a committee weeding out requests for help with the formation of networks in fields as different as "polar science" and "longitudinal studies of development". Evidently, ESF has been embarrassed by the good proposals which have come its way. Only nine of thirty or so have been accepted so far.

Some of the networks seem to have fired the enthusiasm of their participants. Thus the West German participants in the polar science network are eagerly planning to carry several fellow-members on a largely biological exploration of the Antarctic pack-ice next Antarctic spring and summer, using substantial resources that never touch ESF's hands in the process. Much the same is true of the network of "molecular neurobiology of mental illness", which is really a hunt for European families likely, on psychiatric grounds, to provide genetic polymorphisms likely to be technically informative of the genetic basis of psychiatric illness.

The success of the networks is both an embarrassment and a pointer to ESF's general difficulties. Burgen's committee has been aiming to find four new networks a year. Its strategy is to nurse each of them through the first two years (Phase I) and then to persuade member research councils to keep them going longer. The direct cost of the first two years is estimated at FF 550,000. By "passing round the hat", ESF has collected a seed-fund of nearly FF 10 million, but most of that is already committed. That is why a Norwegian delegate at last week's meeting pleaded for restraint — perhaps two and not four networks a year. Evidently, the substrates are beginning to fear what the catalysts may provoke financially.

But if ESF has struck a vein of European collaboration that excites the respect of working researchers, should not the argument be turned around? If an activity which is generally welcomed, not merely by the participants but by organizations such as the European Commission, would it not make sense to increase the funds available? And should not ESF have enough cash to be able to hold what are exploratory meetings such as those planned for its environmental toxicology project? The general opinion is that ESF's budget should be ten times bigger than it is.

Where could the money come from? National research councils fear that more for ESF would mean less for them, but there is a case that governments should contribute something through their foreign offices. So should the European Commission. Luckily, for ESF, the networks should so quickly make ESF much better known that there may soon be an international lobby of researchers for the continuation and growth of its good works.

John Maddox

Superconducting ceramics

Solitons and superconductivity

Alan Bishop

THE discovery of the new ceramic superconductors, whose high transition temperatures cannot easily be explained by traditional ('BSC') theories of superconductivity, has provoked many theoreticians to dust off and remould exotic theories originally intended for other circumstances. Even though the contributions are coming from very different fields of physics, that of T. D. Lee (who is best known for his theory of parity violation in nuclear disintegration) on page 460 of this issue¹ may seem more surprising than most. He shows that several theories invoke a single mathematical concept — the soliton² — that itself has appeared in many diverse disciplines.

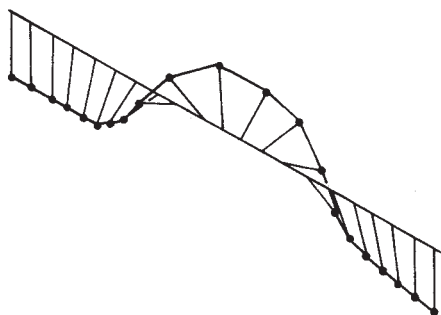
In a strict mathematical sense, solitons are solutions of a limited subset of non-linear partial differential equations in one space dimension and time. Examples are the sine-Gordon, nonlinear Schrödinger and Korteweg-de Vries equations, approximately describing phenomena in hydrodynamics, optics, plasmas and much more². The most familiar solitons have spatially limited kink, pulse or envelope shapes (see figure). Solitons interact with each other completely nondestructively, suffering only phase-shifts during collisions. This ideal particle-like characteristic is one of many remarkable properties accompanying the integrable structure of the underlying equations.

The notion of a soliton has been widely generalized in the past decade, especially by physicists in condensed-matter and field theory. In practical usage², it now signifies a particle-like lump — a spatially localized, finite-energy collective structure that may or may not propagate and that need not suffer purely elastic collisions with other solitons. In this generalized sense, soliton ('defect') classification in higher spatial dimensions by topological arguments has been valuable, such as in nonlinear elastic media, superfluid helium and field theory (where solitons have been advocated as the means to unify particles and field equations).

A similar situation prevails in discussions of quantum solitons². Quantization of any soliton is possible in principle but exact procedures are available for strict solitons only. In fact, quantizing strict solitons is precisely equivalent to solving 'exactly solvable' problems in one-dimensional quantum many-body physics or field theories, or two-dimensional classical statistical mechanics.

An important example of a generalized soliton in condensed matter² is the polaron, often referred to as a 'bag' in

equivalent field-theoretical problems. Coupling between two or more fields results in a self-consistent feedback that becomes stabilized as a localized deformation in the fields. For example, coupling between electrons (or their absences, termed holes) and phonons can result in localization of an electron in a localized deformation of the phonon field, that is of the associated atomic lattice — the elastic energy lost in deforming the lattice is compensated by a gain in electronic energy. Describing the structure and dynamics of polarons has exercised the best theoreticians for several decades as a fundamental problem in many-body physics³. But these objects have enjoyed a revival with the



A mechanical model of a soliton. Rods can swing around a rigid bar and are connected by elastic string. A twist, once introduced, cannot be removed, and keeps its shape because of tension in the string. (Courtesy of Ian Stewart and Oxford University Press.)

advent of quasi-one- and two-dimensional solid-state materials because they are stable over larger parameter ranges in these reduced dimensions.

There are several important characteristics of the remarkable new high-temperature superconductors. First, they are ceramics, not metals, so it is not surprising that they display microtexture on many scales. Second, it seems that the active components are planes (or possibly chains) of copper and oxygen atoms. The reduced dimensionality enhances competition between different modes of collective ground-state ordering of spins and/or charges of the copper ions. Such competitions can be described within familiar model hamiltonians (such as Hubbard models) but the interplay of low-dimensionality and the coupling strengths of electron-electron and electron-phonon interactions forces us to address unfamiliar regimes of those models. Third, excitations created by changing the stoichiometry or the doping (counter) ions (which result in adding electrons or holes in the ground state) are similarly rich, depending on the reference ground state as well as

the electron-electron and electron-phonon interactions.

In this setting, solitons are becoming a recurrent theme in the rapidly evolving theoretical literature. First, microstructures observed by transmission electron microscopy show 'twinning' textures — 'solitonic' striped phase lattices of dislocations (with a spacing of several hundred Ångströms) separating variants of the low-temperature lattice structure. Such microstructure can affect the mechanical strength of the materials. Also, it is hotly debated^{4,5} whether the microstructure can substantially raise the transition temperature of the superconducting phases. Second, the excitations referred to above in simple many-body hamiltonians are typically polaron-like localized states. Depending on whether it is purported that the superconductivity is produced predominantly by spin fluctuations, charge fluctuations or both, theories lead to electron (or hole) polarons, spin-polarons (spin fluctuations localized in a lattice deformation) or similar 'spin-bags'. A popular notion is then that these polarons may bind in pairs (bipolarons) which can behave as 'bosons' and undergo Bose condensation into a macroscopic superconducting coherent quantum state. Relationships between real space pairing and the usual momentum space pairing of the conventional BCS theory of superconductivity also arise in polaron-based descriptions of strong-coupling BCS.

Whether any of the above soliton mechanisms are relevant in the new materials is still not clear. Indeed it is quite possible that different mechanisms will operate as we learn to tune the synthesis of the materials through a more general class. T. D. Lee's contribution in this issue emphasizes from a field-theoretical perspective the polaron nature of many of the proposed many-body approaches. He observes that the various theories typically lead to equations studied elsewhere (often as model field theories) and known to have soliton 'bag' solutions. A central feature of soliton equations is indeed their genericity — the same equation can arise in very different physical contexts because of their common underlying ingredients. With so many new researchers attracted by the challenge of high-temperature superconductors, it is important that the lesson from soliton studies of cross-referencing between different disciplines is not lost. □

1. Lee, T. D. *Nature* **330**, 460–461 (1987).
2. Bishop, A. R. *et al. Nonlinearity in Condensed Matter* (Springer, Heidelberg, 1987).
3. Emin, D. *Physics Today* **34** (June 1982).
4. Horowitz, B., Krumhansl, J. A. & Barsch, G. *Phys. Rev. B* (in the press).
5. Mueller, F. *et al. Phys. Rev. Lett.* (in the press).

Alan Bishop is in the Theoretical Division and the Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA.

Palaeontology

Britain's newest mammoths

Jeffrey J. Saunders

FOR many years those who borrowed from palaeontological data believed that mammoths disappeared from Britain during the maximum expansion of ice sheets during the Last Cold Stage, between 18,000 and 15,000 years before present (BP), and that they did not reappear there during the subsequent ameliorated climates of the late-glacial. This absence was thought to result from generally reduced mammoth populations on the European continent following that last glacial maximum. On page 472 of this issue¹, Coope and Lister report the discovery of mammoth skeletons in late-glacial kettle deposits at Conover, Shropshire, England. Their reported radiocarbon dates of $12,920 \pm 390$ BP and $12,700 \pm 160$ BP are the youngest for mammoths in Britain, averaging approximately 5,000 years younger than those previously available. It appears, furthermore, that all was well with these mammoths in their late-glacial British habitat, and it is no surprise that this new unexpected record is an important and interesting discovery.

The Last Cold Stage in Britain is the Devensian, extending from approximately 110,000 to 10,000 BP². Although climatically complex, it was predominantly characterized by a cold climate that favoured permafrost conditions and subarctic park-like landscapes. During the Devensian glacial maximum 18,000–15,000 BP, with sea water locked in glacier ice and sea level perhaps as much as 100 m lower than today, the unglaciated southern portions of England and Wales were joined to the European continent as part of a broad contiguous region of predominantly tundra, north of the Alps and extending east to Siberia. The Devensian

was succeeded 10,000 years ago by the Flandrian, the current interglacial characterized by temperate deciduous forest. The Late Devensian vertebrate faunal history of Britain² is recorded at 11 primary localities in England, from the Isle of Man and York southwards, and at Ballybetagh^{2,3} in Ireland. To these Conover can now be added.

Mammoth (*Mammuthus primigenius*), horse (*Equus ferus*), woolly rhinoceros (*Coelodonta antiquitatis*) and reindeer

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Cold comfort — did mammoths survive the Last Cold Stage?

(*Rangifer tarandus*) occur in the earlier Late Devensian assemblages (about 26,000–18,000 BP) in association with tundra-like vegetation and arctic climates. Artefacts are not associated with these faunas. Until the discovery at Conover, later Late Devensian faunas lacked mammoths, but horses, giant deer (*Megaloceros giganteus*), European elk (*Alces alces*) and reindeer are recorded associated with birch woodlands or parklands containing grasses. Artefacts occur in association with these faunas after 12,000 BP.

One of the implications in Coope and Lister's report is particularly noteworthy.

This is the issue of large-mammal extinction during the late-glacial. Mammoths later than the Conover record are unknown in Britain, where later late-glacial fossil deposits also lack woolly rhinoceroses². Although discoveries such as that at Conover leave open the issue of timing, mammoths, woolly rhinoceroses and giant deer are extinct today, part of a pattern of extinction throughout the Northern Hemisphere (and elsewhere) that was complete by 10,000 BP. Cultural as well as climatic causes have been proposed for this event.

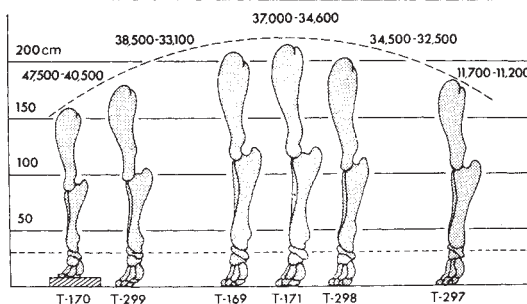
Did the last mammoths of Britain die out before those of America? Coope and Lister¹ have reduced the temporal differences from 7,000 years to about 2,300 years (12,800 BP at Conover versus

10,500 BP at Rawhide Butte, Wyoming, the youngest date for mammoths in America⁴). As additional well-dated late-glacial mammoths are found in Britain and on the continent (now a reasonable expectation), temporal differences in late or last occurrences may be reduced. If these temporal differences could be reduced to zero, then late-arriving native Americans at 12,000 BP would seem less likely to be sole agents in the extinction of the American megafauna⁵. Events on a global scale would be sought and the most likely of these would be climatic change. Global glacial recession during the Last Cold Stage seems to have been harmonious (Devensian², Weichselian⁶, Valdaian⁷, Wisconsinian⁸) and indicates parallel patterns of late-glacial climatic change. This change seems to have been rapid and to have occurred through two phases of climatic amelioration: the first commencing 13,000 years ago, and interrupted by climatic deterioration at 11,000 BP, in turn followed by the second abrupt amelioration about 10,000 BP, resulting in the current interglacial. In America (and perhaps elsewhere) late-glacial extinctions occurred during the climatic deterioration at 11,000–10,000 BP and therefore correspond with the end of the Last Cold Stage. □

sion during the Last Cold Stage seems to have been harmonious (Devensian², Weichselian⁶, Valdaian⁷, Wisconsinian⁸) and indicates parallel patterns of late-glacial climatic change. This change seems to have been rapid and to have occurred through two phases of climatic amelioration: the first commencing 13,000 years ago, and interrupted by climatic deterioration at 11,000 BP, in turn followed by the second abrupt amelioration about 10,000 BP, resulting in the current interglacial. In America (and perhaps elsewhere) late-glacial extinctions occurred during the climatic deterioration at 11,000–10,000 BP and therefore correspond with the end of the Last Cold Stage. □

Size change in Siberian mammoths (*M. primigenius*) during the Last Cold Stage. Size of fore-limb is compared to radiocarbon age. In Siberia the large forms date to the Karginsk interstadial where their size is thought to be a response to optimal foraging conditions. The environment indicated for the Conover mammoths is temperate to boreal, but not arctic, and a "relatively mild climate is implied" (ref. 1).

Radiocarbon dates indicate an early Windermere Interstadial age for the Conover remains and, based on the Siberian model, the early part of this late-glacial warming of Britain was accompanied by the return of optimal mammoth habitat. (From ref. 9, after ref. 10).



1. Coope, G.R. & Lister, A.M. *Nature* **330**, 472–474 (1987).
2. Stuart, A.J. *Pleistocene Vertebrates in the British Isles* (Longman, London, 1982).
3. Barnosky, A.D. *Science* **228**, 340–344 (1985).
4. Meltzer, D.J. & Mead, J.I. *Quat. Res.* **19**, 130–135 (1983).
5. Martin, P.S. in *Quaternary Extinctions* (eds Martin, P.S. & Klein, R.G.) 354–403 (Univ. Arizona Press, Tucson, 1984).
6. Björck, S. & Möller, P. *Quat. Res.* **28**, 1–37 (1987).
7. Soffer, O. *The Upper Paleolithic of the Central Russian Plain* (Academic, Orlando, 1985).
8. Mott, R.J. et al. *Nature* **323**, 247–250 (1986).
9. Kurtén, B. *Pleistocene Mammals of Europe* (Aldine, Chicago, 1968).
10. Heintz, A.E. & Garutt, V.E. *Norsk geol. Tidssk.* **45**, 73–79 (1965).

Jeffrey J. Saunders is at the Quaternary Studies Center, Illinois State Museum, Springfield, Illinois 62706, USA.

Retinoic acid receptor

Towards a biochemistry of morphogenesis

Miranda Robertson

It is axiomatic that the development of multicellular organisms depends upon morphogenetic signals between cells that instruct changes in gene expression. More particularly, concentration gradients of chemical morphogens are believed to act across small fields of embryonic cells where different concentrations of morphogen provide 'positional information' that determines the future behaviour of the cells. Until recently however, while the logical case for such mechanisms was impeccable, the biochemical evidence was inferential at best. The position has now changed. Five months ago, Thaller and Eichele¹ reported evidence strongly implicating gradients of retinoic acid in the specification of the anterior-posterior digit pattern of the chick wing (see ref. 2); and in this and forthcoming issues Petkovich *et al.*³ and Giguere *et al.*⁴ announce the identification of the retinoic acid receptor (RAR) as a member of the steroid/thyroid receptor family.

This family of intracellular receptors consists of a group of proteins that mediate responses to largely lipid-soluble ligands by binding to specific regulatory sequences in the DNA and directly bringing about changes in gene expression. Thus not only has a vertebrate morphogen been identified for the first time, but its mechanism of action also seems clear, in principle if not in detail.

Two papers to appear in forthcoming issues of *Nature*^{4,5} show that there are other members of this family whose ligands have yet to be identified, tempting speculation that some of them may be receptors for unidentified morphogens.

With other lines of evidence implicating members of the same family in oncogenesis^{5,6}, the only family that seems likely to attract a comparable level of intrusive investigation in 1988 is the Royal House of Windsor.

Family relationships

The steroid/thyroid receptor family comprises some eight identified proteins. These proteins all bind signal molecules that diffuse through the plasma membrane and induce a change in the receptor that enables it to bind in turn to regulatory sequences in the cellular DNA. The mechanism of this ligand-dependent DNA binding is unknown⁶.

Analysis of the gene family encoding the receptor proteins has shown that they fall broadly into two groups: one comprises the receptors for the steroids

oestrogen, glucocorticoid, progesterone⁶, aldosterone⁷ and vitamin D⁸ and the other consists of six or seven genes two of which are known to encode thyroid receptors (TR α and TR β)⁹⁻¹¹, while two others were detected by hybridization to sequences on chromosome 17 (ref. 3), and one came to light because of its activation by the integration of hepatitis B virus (HBV) DNA in a hepatocellular carcinoma¹². All members of the family are related to the viral oncogene *erb-A*¹³, the thyroid receptors more closely than the steroid receptors; indeed TR α is encoded by the cellular *erb-A1* gene. It is to this group that the retinoic acid receptor belongs.

Finger swap

The identification of RAR was made possible by the ingenious exploitation of homologies in two different ways. All genes in the steroid/thyroid receptor family can be divided into four regions, two of which, designated C and E, encode functional domains mediating DNA and ligand binding, respectively (see figure). The most highly conserved region is C which encodes the DNA-binding domain, and which has the sequence motif characteristic of the DNA-binding 'finger' domains first identified in the TFIID transcriptional activator (see ref. 6). Petkovich *et al.*³ used hybridization to a consensus C-region sequence to isolate the RAR gene, and Giguere *et al.*⁴ used hybridization to the sequence identified by Dejean *et al.*¹² at the site of HBV integration. Both used expression vectors to produce the RAR protein (otherwise

not an abundant species) in large enough quantities to show that it binds retinoic acid with high affinity.

A direct functional assay for transcriptional regulation by the occupied RAR is not possible, because the DNA sequences conferring retinoic acid responsiveness have not been identified. However because of the homologous modular organization of all the nuclear receptor genes (see figure), it was possible to circumvent this difficulty by a manoeuvre colloquially known as the finger swap, in which the DNA-binding (C) region of one receptor is exchanged for that of another.

In this case, the C region of the RAR cDNA was replaced with the corresponding C region of the oestrogen-receptor³ or the glucocorticoid-receptor gene⁴. Cultured cells were then co-transfected with this chimaeric construct in a suitable expression vector, and a second construct consisting of a reporter gene (the bacterial chloramphenicol acetyltransferase (CAT) gene) linked to an appropriate regulatory element — the oestrogen-responsive element in the case of Petkovich *et al.*, and the mouse mammary tumour virus long terminal repeat in the case of Giguere *et al.*. This provided a functional assay for transcriptional activation, and both groups were able to show that retinoic acid at concentrations known to be effective *in vivo* would activate transcription of the CAT gene.

Black sheep and dark horses

Although it is the morphogenetic effects of retinoic acid whose implications are furthest-reaching, it has other effects that are significant in themselves and may be important in suggesting approaches to some of the questions that arise from the identification of the receptor. As well as playing a part in the growth and maintenance of epithelia, retinoic acid induces differentiation in a number of cell types *in vitro*. For example, F9 teratocarcinoma cells can be induced by retinoic acid to differentiate into parietal endoderm, with the appearance of specific mRNAs. This may make it possible to identify specific DNA binding sites for the RAR.

At the same time, finger-swap assays may make it possible to identify ligands for the products of those receptor genes whose functions are still unknown. These now number at least four: two *erb-A*-related genes on chromosome 17; the HBV-activated gene, now designated *hap*²; and at least one more gene that hybridizes with the RAR sequence on Southern blots³ (see table).

(It should be noted that there is one major discrepancy in the RAR sequence reported by Petkovich *et al.* and by Giguere *et al.*, in that the Giguere *et al.* initiation codon is 30 codons further upstream than that of Petkovich *et al.*. While none of the



Schematic diagram of a generalized steroid/thyroid receptor gene, showing the division of the gene into regions A/B, C, D and E. The function of region A/B is unknown; C encodes the DNA-binding domain; D is believed to be a hinge region, and E encodes the ligand-binding domain. In general, the DNA-binding domain is most highly conserved, both within and between the two groups of receptor (steroid and thyroid); the ligand-binding domains show less homology. With the addition of data from the retinoic acid receptor, it seems that the two domains have evolved independently in the two families⁴, suggesting evolution by exon shuffling. It is possible through exon-shuffling by recombinant DNA technology to produce hybrid receptor molecules for use in functional assays (see text).

cDNAs isolated by Petkovich *et al.* contains the upstream AUG identified by Giguere *et al.* (although they are long enough), genomic DNA isolated in the Strasbourg Laboratory precisely corresponds to the sequence they report (P. Chambon, personal communication). The reasons for the divergence of the two cDNA sequences within the first 30 codons are not clear; one possibility is a cloning artefact in the Chambon laboratory. The discrepancy falls within the A/B region and does not affect the conclusions, nor is there any doubt that both laboratories have the same gene, the Giguere *et al.* sequence almost certainly being the complete one.)

There are three obvious ligands still looking for receptors: testosterone, dioxin and retinol; so with a minimum of four receptors looking for ligands, at least one new hormone, possibly also a morphogen, is to be expected.

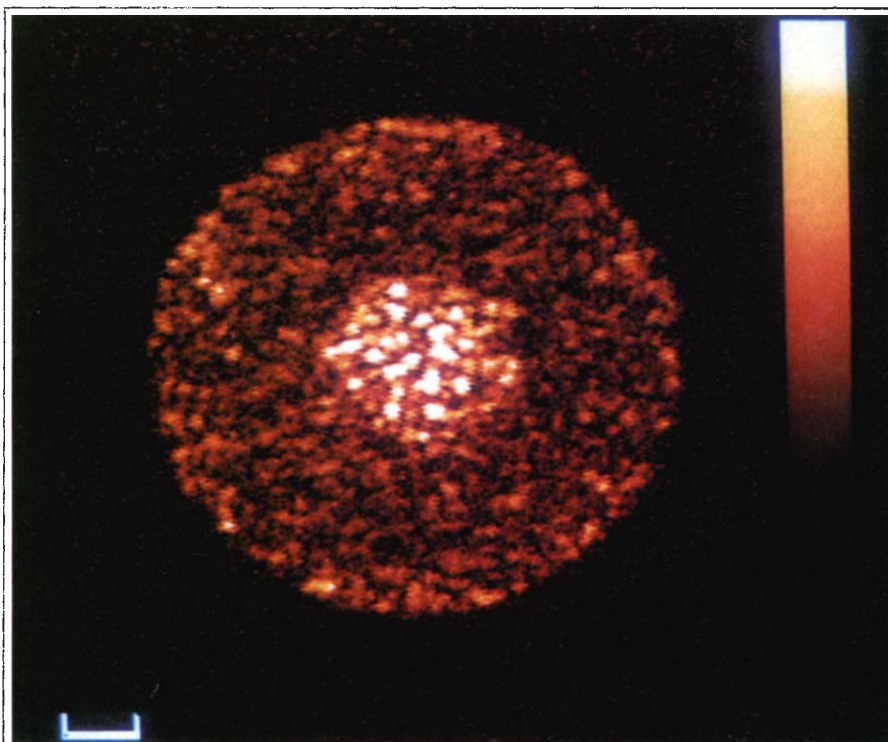
In the meantime, the *hap* gene, whatever its ligand, seems to illustrate the principle that neoplastic transformation may occur through inappropriate expression of a regulatory gene: of two *hap* transcripts presumably generated by differential splicing, only one is detected by de The *et al.*⁵ in normal liver, but both appear in hepatoma cells.

Ontogeny and phylogeny

The discovery, in the RAR, of a putative morphogen receptor offers for the first time in a vertebrate system the hope of analysing the mechanism of morphogenesis by identifying a set of developmental control genes.

Two obvious questions to ask of the chick limb are, first, what is the pattern of expression of the RAR in the limb bud; and second, what are the genes whose activities are changed in response to retinoic acid in this tissue?

It is perhaps of particular interest that the gradient of retinoic acid across the chick limb bud is relatively shallow (~50 nM at the anterior edge to ~20 nM at the posterior edge). With a K_d of between 10^{-9} and 10^{-8} for the association of retinoic acid with its receptor, this would mean an expected change of only about 15 per cent in maximum binding from thumb to little finger. It might be argued (P. Chambon, personal communication) that this is a rather small difference to supply positional information adequate for the specification of these structures, and some amplification mechanism should be



ULTRASOUND backscatter microscopy reveals the internal structure of a living tumour spheroid, 850 μ m in diameter, derived from a human bladder carcinoma cell line. The viable cells in the rim of the spheroid surround a denser necrotic core. This new technique, at the moment capable of providing tomographic images to a depth of 4 mm, is described by M.D. Sherar, M.B. Noss and F.S. Foster of the Ontario Cancer Institute on page 493 of this issue. Scale bar, 100 μ m.

sought. One possibility for such an amplifying mechanism would be a gradient of the receptor itself; a second might involve some interaction with the cytoplasmic retinoic acid binding protein (CRABP)³; and a third is cooperative binding of the occupied RAR to the DNA. Multiple DNA binding sites are already known for some steroid receptors, and there is some functional evidence for binding cooperativity¹⁴. The advantage of such a mechanism in the context of morphogenesis would be that it could provide a mechanism for the kind of threshold response needed to account for the specification of discontinuous structures by a continuous gradient of signal.

In an article published in 1975, Gordon Tomkins speculated¹⁵ on how the morphogenetic signals that initiated the evolution of multicellular organisms might have arisen from the signals whereby in unicellular organisms information about nutrient availability was conveyed to the genes encoding the

relevant biosynthetic enzymes. From this perspective, the long-range action of hormones would have been the latest evolutionary function of the steroid/thyroid receptor signalling mechanism, which could have evolved from a primitive system enabling positional information to be conveyed by hydrophobic molecules diffusing from cell to cell, perhaps as a modification of an existing metabolic control system. Some metabolically controlled genes in yeast have recently been shown to have multiple upstream binding sites to which regulatory proteins bind cooperatively^{16,17}.

Steroid/thyroid receptor gene family

Steroid receptors	Non-steroid receptors	Unassigned genes
Oestrogen (6)	Thyroid hormone	<i>Erb-A2</i> -hybridizing (17q21.3)
Glucocorticoid (5)	(α -receptor) (17q11.2-q12)	<i>Erb-A2</i> -hybridizing (weak)
Progestin (11q23)	Thyroid hormone	(17q25)
Mineralocorticoids (4)	(β -receptor) (3)	<i>hap</i> (3)
Vitamin D	Retinoic acid (17q21.1)	RAR-hybridizing

The numbers in parentheses are the chromosomal localizations where known.

- Thaller, C. & Eichele, G. *Nature* **327**, 625-628 (1987).
- Slack, J.M.W. *Nature* **327**, 553-554 (1987).
- Petkovich, M., Brand, N.J., Krust, A. & Chambon, P. *Nature* **330**, 444-450 (1987).
- Giguere, V., Ong, E.S., Segui, P. & Evans, R.M. *Nature* (in the press).
- de The, H., Marchio, A., Tiollais, P. & Dejean, A. *Nature* (in the press).
- Green, S. & Chambon, P. *Nature* **324**, 615-617 (1986).
- Arriza, J.L. *et al.* *Science* **227**, 268-275 (1987).
- McDonnell, D.P. *et al.* *Science* **235**, 1214-1217 (1987).
- Sap, J. *et al.* *Nature* **324**, 635-640 (1986).
- Weinberger, C. *et al.* *Nature* **324**, 641-646 (1986).
- Thompson, C.C., Weinberger, C., Lebo, R. & Evans, R.M. *Science* **237**, 1610-1614 (1987).
- Dejean, A., Bougleret, L., Greshchik, K.-H. & Tiollais, P. *Nature* **322**, 70-72 (1986).
- Bishop, M. *Nature* **321**, 112-113 (1987).
- Jantzen, H.-M. *et al.* *Cell* **49**, 29-38 (1987).
- Tomkins, G.M. *Science* **189**, 760-763 (1975).
- Giniger, E., Varnum, S. & Ptashne, M. *Cell* **40**, 767-774 (1985).
- Giniger, E. & Ptashne, M. *Proc. natn. Acad. Sci. U.S.A.* (in the press).

Miranda Robertson is Biology Editor of Nature.

Amorphous silicates

Uneconomic bonding in glasses

Bernard de Jong

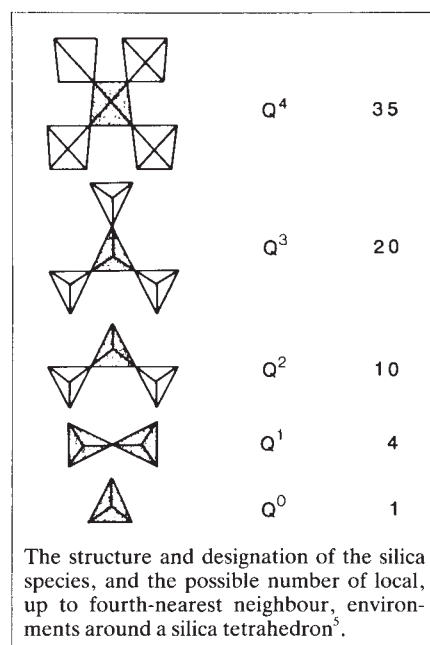
STUDIES of amorphous materials, those having no form or (by implication) structure, have become increasingly sophisticated over the past ten years. Geologists wishing to understand the crystallization sequence of molten aluminosilicates have resorted to model studies of simple systems. Such a study, on amorphous sodium silicate, is reported by J.F. Stebbins on page 465 of this issue¹. Stebbins resolves an important argument about the types of bonding in glasses, showing that more types are present than are required stoichiometrically.

New ideas about the structure of amorphous silicates arose in the 1970s as a result of experimental evidence for alkali clustering and for the presence of a range of silica species. It had long been recognized that the liquidus curve in alkali-silicate phase diagrams deviates from the ideal shape, most strongly in the lithium-silicate system². Standard thermodynamic theory explained such deviation as being caused by clustering of the lithium atoms. It remained an open question how such clustering came about and if there existed any experimental evidence for it on an atomic scale. A partial answer to this question required characterization of the types of oxygen in a glass. At most three types can occur: bridging oxygens associated with two silicon atoms, Si—O—Si; non-bridging oxygens associated with a silicon and alkali atom (R), Si—O—R; and non-bonding oxygens associated with alkalis only, R—O—R.

In the first experimental study on the concentration of these three different types of oxygen atoms in silicate glasses, Brückner and co-workers showed that the shape of the oxygen 1s spectral line indicates that non-bonding oxygens are not present in silica-rich alkali-silicate glasses³. Hence, clustering of alkalis does not create non-bonding oxygens and by implication involves linkage to a silicon atom. Molecular orbital calculations for the canonical $H_6Si_2O_7$ molecule show that clustering of alkalis in an amorphous silicate is reflected by preferential attachment of the lighter alkalis to silicon tetrahedra that already have an alkali attached to them⁴. These results suggested that, within stoichiometric constraints, four lithium atoms would be attached to a single silica tetrahedron in amorphous lithium silicates, whereas such a configuration would not occur in equimolar amorphous caesium silicates.

The second line of evidence came from Fourier-transform nuclear magnetic resonance (NMR) spectra of aqueous

alkali-silicate solutions^{5,6}. Spectra of such solutions show the presence of a distribution of different silica species that varies as a function of hydroxyl content, confirming previous potentiometric results^{7,8}. Five species were discovered, designated Q^0 – Q^4 , indicating that the central silicon-oxygen tetrahedron is surrounded by 0–4 bridging oxygens (see figure). Moreover, the character of the third-nearest-neighbour silicon atom affects the ^{29}Si NMR chemical shift. The number of possible variations of local environment was



evaluated taking such third-nearest-neighbour interactions into account, as shown in the figure.

Our experiments in the early 1980s to determine the silica species distribution in amorphous alkali silicates⁴ used the variations in line shape of the silicon $K\beta$ X-ray emission spectrum to determine $Q^4/(Q^0+Q^1+Q^2+Q^3)$ (or Q^4/Q^R) ratios. In addition we found that different alkalis in alkali-silicate glasses give different Q^4/Q^R ratios. Individual species distributions, however, were still not amenable to experimental verification. It was hoped that ^{29}Si magic-angle-spinning NMR⁹ would change that, but poor spectral resolution remained a problem. Even in favourable cases, only Q^4/Q^R ratios could be determined unequivocally¹⁰. Spectral assignments to different species were tentative because silicon chemical shifts in the glasses do not correspond one-to-one with those in crystalline precipitates. Also the species distributions that were determined using best-fit gaussians did not give the

proper stoichiometry in the high-silica glasses. Two schools of thought developed: one maintaining that only one or two species are present and hence suggesting an economy in the distribution¹⁰; the other that all species were present in varying concentrations which would be tantamount to an uneconomic distribution^{11,12}.

The paper by Stebbins in this issue¹ addresses one main question and three secondary ones. The main question is that there is a reasonably quantifiable, uneconomic species distribution, consisting of Q^2 , Q^3 and Q^4 , in the composition range considered. This coincides with our conclusion based on NMR results that no single silica species exists in aqueous silicate solutions even in dilute (100 parts per million) solutions¹³. The secondary results are that the NMR lines are not as broad as anticipated, indicating that sodium-silicate glasses are not as aleatoric as expected from the number of possible configurations shown in the figure. Furthermore, Stebbins' results suggest that sodium-silicate glasses tend towards a narrower species distribution than that encountered in lithium-silicate glasses, demonstrating the difference in impact of different alkalis on the structure of alkali-silicate glasses. Finally, his results indicate that radial-distribution analysis samples the character of the average structure in contrast with ^{29}Si magic-angle-spinning NMR, which samples the distributive structure. Taken together, Stebbins' results show that sodium-silicate glasses are weakly random relative to glass systems with a binomial distribution of silica species.

Stebbins' convincing proof for the existence of a silica species distribution and its variation with different alkalis is an important step in unravelling the structure of amorphous alkali silicates. The principal remaining issue in explaining the physical properties of amorphous alkali silicates is how these different species are connected, and also how this connectivity varies as a function of the type of alkali present and what its effect is on the percolation threshold. □

1. Stebbins, J.F. *Nature* **330**, 465–467 (1987).
2. Kraceck, F.C. *J. phys. Chem.* **34**, 2645–2665 (1930).
3. Brückner, R., Chun, H.-U. & Goretzki, H. *Glastech. Ber.* **49**, 211–213 (1976).
4. de Jong, B.H.W.S., Keefer, K.D., Brown, G.E. & Taylor, C.M. *Geochim. cosmochim. Acta* **45**, 1291–1309 (1981).
5. Engelhardt, G., Zeigler, D., Janke, H., Hoebbel, D. & Weiker, W. *Z. anorg. allg. Chem.* **418**, 17–28 (1975).
6. Marsmann, H.C. *Z. Naturf.* **B29**, 495–459 (1974).
7. Lagerström, G. *Acta chem. Scand.* **33**, 722–736 (1959).
8. Ingri, N. *Acta chem. Scand.* **13**, 758–775, (1959).
9. Lippmaa, E. et al. *J. Am. chem. Soc.* **102**, 4889–4893 (1980).
10. de Jong, B.H.W.S., Schramm, C.M. & Parziale, V.E. *J. Am. chem. Soc.* **106**, 4396–4402 (1984).
11. Grimmer, A.R. et al. *Phys. Chem. Glasses* **25**, 105–109 (1984).
12. Dupree, R., Holland, D., McMillan, P.W. & Pettifer, R.F. *J. Non-Cryst. Solids* **68**, 399–410 (1986).
13. Cary, L.W., de Jong, B.H.W.S. & Dibble, W.E. *Geochim. cosmochim. Acta* **46**, 1317–1320 (1982).

Bernard de Jong is in the Department of Applied Physics, Corning Glass Works, Corning, New York 14831, USA.

Conservation biology

Justifiable killing of birds?

Jared M. Diamond

A BITTER debate has been triggered in Australia by the recent death of a leading ornithologist, Julian Ford, after his arrest for alleged illegal collecting of birds. The affair provokes the question of whether scientific collecting of birds is at all justifiable today, when wholesale destruction of habitats for timber and agriculture is already drastically reducing populations.

One answer is that all decisions about conservation, wildlife regulations and creation of new national parks are based on faunal and floral catalogues defined by the information that specimens provide about species and races, their geographical variation and distribution. For instance, collectors in Queensland recently described five new bird species (see figure), several of them very local and endangered, but all so similar morphologically to widespread related species that they would never have been established as separate species without comparisons of series of specimens. Only when these species had been recognized could conservation measures for their protection be initiated.

The other answer is that specimens are essential for understanding many basic questions of biology. No one denies that collecting can be misdirected or that it should be regulated. But much of our biological knowledge still rests on evidence from specimens, particularly of birds, which have contributed uniquely to our understanding of evolution. It is no accident that Galapagos finches were central to Darwin's grasp of evolution, and that only when his specimens had been brought to England did he realize that the finches comprised many species whose origin required explanation. Those specimens that Darwin shot continue to be important to modern biologists (Lack, D. *Darwin's Finches*, Cambridge Univ. Press; 1947; and Grant, P.R. *Ecology and Evolution of Darwin's Finches*, Princeton Univ. Press; 1986).

It is often assumed that Darwin solved the main problems of evolutionary biology, and that collection of more specimens is unnecessary. But evolution is full of unsolved questions. What role do peripheral isolates play in speciation? Does selection against hybrids sharpen reproductive isolation? Do ecological niche differences appear before or after reproductive isolation? Adequate series of bird specimens from carefully chosen localities provide material for addressing these questions. Australian birds are particularly important because past cycles of rain-fall in that country caused populations to expand and contract and hence become

isolated, re-encounter each other and hybridize or diverge. Julian Ford excelled at identifying critical isolates and hybrids, and at combining statistical analyses of bird morphology with field observations of bird behaviour. He wrote classic papers on the evolution of quail-thrushes and other bird groups, and on the origin of Australia's distinctive desert and



Female *Eclectus parrot*, confined in Australia to Queensland's rain forests. The emerald-green male looks so different from the female that they were thought to be of different species.

mangrove avifaunas (*Emu* 74, 161–186, 1974; 82, 12–23, 1982; 83, 152–172, 1983; and 86, 87–110, 1986).

To place Ford's death in perspective, consider the following facts. On the one hand, he was charged with collecting 600 bird specimens, but on the other, a square mile of rain forest supports around 10,000 birds, most of them belonging to species that cannot survive outside rain forest and are doomed if their habitat is destroyed. Yet governments that stringently police collecting not only fail to stop the felling of trees, but grant logging concessions on enormous tracts of land, thereby trading short-term economic gain for long-term loss, as the areas of rain forest throughout the world are shrinking rapidly. Why such zeal to prevent a few birds from being collected for science, while killing millions of birds without contributing to knowledge?

There are four reasons why, throughout the world, governments indifferent to conservation like to be tough on scientific collecting: first, it takes thought to realize that habitat destruction kills birds as surely as do guns; second, ornithologists are few, impoverished and politically impotent compared with large industries; third, the reasons for scientific collecting make duller newspaper reading than do the arguments of animal-rights lobbies; and finally, harassing scientists offers a cheap way to feign concern for conservation. If authorities really want their conservationist zeal to be efficacious, they could begin by protecting their natural habitats. □

Jared M. Diamond is Professor of Physiology at the University of California Medical School, Los Angeles, California 90024, USA.

Supernova 1987A

Hard X-rays imply more to come

Donald D. Clayton

THE recent detection of hard X-radiation from the supernova 1987A, reported and discussed by several groups in *Nature* on 19 November^{1–5}, raises expectations that γ -ray lines from radioactive ^{56}Co may soon be detected. Low-energy (soft) X-rays have been detected from other supernovae (although not³ from SN1987A), but hard X-rays and γ -rays from supernovae have never been observed. Astrophysicists have hoped that they will be seen, so that the details will help explain the origin of the elements.

Hoyle initiated the theory of nucleosynthesis in stars in the decade following the Second World War. It was a correction during 1967–68 of an important detail of his initiating paper⁶, on the origin of the iron abundance peak (the equilibrium or e-process), that led to the conviction that the numerous $^{56,57}\text{Fe}$ nuclei were ejected explosively from stars as $^{56,57}\text{Ni}$ radioactive

progenitors rather than as the stable daughters^{7,8}. The following year we realized⁹ that γ -rays emitted during the radioactive sequence $^{56}\text{Ni} \rightarrow ^{56}\text{Co} \rightarrow ^{56}\text{Fe}$ should be detectable. But those of us involved held faint hope of an appropriate supernova in our lifetimes, which makes SN1987A seem a little like Christmas.

The first two positive detections of X-rays emerging from the expanding gases were reported in *Nature* on 19 November by international collaborations using two new space astronomy facilities: the Soviet space station Mir and the Japanese X-ray astronomy satellite Ginga². Ginga, consisting of a set of proportional counters with an effective area of 400 cm² and a field of view of 2° × 4°, made its first detections in early July². The measurements are of importance in that they document the steady growth of the X-ray luminosity in the 10–30-keV band during the summer,

while the optical luminosity was steadily declining. This result is consistent with energetic γ -radiation diffusing from a source by Compton scattering (inelastic scattering by electrons) out of a body growing in surface area and declining in optical depth to its centre.

The energy (E) spectrum seems to be bimodal, a declining thermal spectrum with a temperature of several kiloelectronvolts below $E = 10$ keV, tailing off to a hard flat spectrum in the range 10–30 keV. Two distinct physical mechanisms are clearly indicated. By mid August, two NaI/CsI 'phoswich' detector packages in the Mir–Kvant observatory had detected¹ emission between 20 and 300 keV, with an extremely hard $E^{-1.4}$ power-law spectrum, flattening at low energies to join the Ginga data nicely. Both the high energies of these X-rays and the hardness of their spectrum identify the photons as being the Compton-scattered tails of even higher-energy photons emitted deep within the object.

Indeed, both the X-ray emission and the history of the optical luminosity are well explained by Itoh *et al.*⁴ in the 19 November *Nature*, as resulting from 0.07 solar masses of radioactive ^{56}Ni created in the initial event. Such a yield is expected in the nucleosynthesis accompanying a type II supernova. A series of four papers by Woosley and colleagues (to be published in the *Astrophysical Journal*) also documents this success. The optical light curve was already decaying exponentially with the characteristic lifetime of ^{56}Co ($\tau = 113$ days) when the X-ray power was nearing its expected peak. The X-ray emission too should soon decline, because of the combined effects of decreasing ^{56}Co γ -ray emission and decreasing optical depth for conversion of those γ -rays into X-rays. Because the X-rays are created by the interaction of γ -rays with the atoms and electrons in the supernova envelope, their flux must disappear when the remnant becomes so thin that the γ -rays emerge unimpeded.

How this transition will reveal the structure of the exploding object has also been discussed^{4,11–13}. It is a unique opportunity in astronomy, to know the pre-explosive star and detect an identifiable series of γ -ray lines to document the expanding structure, which, combined with the neutrino burst, will provide vastly more information about the internal structure of a star than has ever been known. Itoh *et al.*⁴ have already concluded from the unexpectedly early maximum in X-ray luminosity that ^{56}Co was mixed outwards via a Rayleigh–Taylor instability thereby reducing its average optical depth. But this conclusion can only be proved by the observation of γ -ray lines themselves.

Detection and identification of the ^{56}Co γ -ray spectrum (and also that from ^{57}Co , which becomes more intense after

3 years^{10,11}) from SN1987A are really needed, not only to confirm that stellar nucleosynthesis has occurred, but also because the γ -ray lines can explain the optical and X-ray light curves, whereas the converse is not true. This is why NASA is supporting five teams that are preparing to launch balloon-borne γ -ray spectrometer packages from Australia (G.J. Fishman, personal communication). NASA's Solar Maximum Mission, already in orbit, is trying to detect γ -ray lines from SN1987A through the collimator walls of its γ -ray telescope (the upper limit in September was $2 \times 10^{-3} \text{ cm}^{-2} \text{ s}^{-1}$). Astronomers hope that the Mir space station will detect γ -rays with its PULSAR phoswich¹. NASA's γ -ray observatory was due to be launched in the delayed 1988 flight of the space shuttle. But a launch in 1990 should suffice to measure the important ^{57}Co lines^{10,11}.

The rapidly declining thermal spectrum² below 10 keV seems likely to arise from hot plasma created when the outermost ejecta collides with circumstellar material. Aschenbach *et al.* searched unsuccessfully for soft X-rays on 24 August using a rocket-launched proportional-counter telescope³. Their result means that the soft X-ray luminosity is more than 2,000 times fainter than that detected from SN1980K in NGC6946, a contrast that they attribute to the very low density of circumstellar matter around the blue-supergiant progenitor, Sanduleak–69 202. This and a parallel discussion by Masai *et al.*⁵, setting the inner radius of circumstellar matter at 10^{16} cm, will reveal much about the strength of stellar winds

from Sk–69 202 and thereby about its evolutionary state before the explosion.

Despite the dramatic success of the radioactivity theory, its correctness is by no means a foregone conclusion. The entire supernova display could instead have been powered by X- and γ -radiation from a rapidly spinning neutron star at the supernova core. Fast young pulsars are known to be strong sources of γ -radiation. Some high-energy γ -ray astronomers hope for this outcome because it will provide them with a bright source of γ -rays in the years to come. Fortunately, the measurements themselves will soon give the answer. A pulsar source would continue to grow in intensity, whereas the radioactivity must decay. Pulsar γ -rays extend to high energy whereas radioactivity emits below 3 MeV. But perhaps we will be given both, as theory predicts. There is something in this incredible event for everyone. □

1. Sunyaev, R. *et al.* *Nature* **330**, 227–229 (1987).
2. Donati, T. *et al.* *Nature* **330**, 230–231 (1987).
3. Aschenbach, B. *et al.* *Nature* **330**, 232–233 (1987).
4. Itoh, M. *et al.* *Nature* **330**, 233–235 (1987).
5. Masai, K. *et al.* *Nature* **330**, 235–236 (1987).
6. Hoyle, F. *Mon. Not. R. astr. Soc.* **106**, 343–383 (1946).
7. Truran, J.W., Arnett, W.D. & Cameron, A.G.W. *Can. J. Phys.* **45**, 2315–2332 (1967).
8. Bodansky, D., Clayton, D.D. & Fowler, W.A. *Phys. Rev. Lett.* **20**, 161–164 (1968).
9. Clayton, D.D., Colgate, S.A. & Fishman, G.J. *Astrophys. J.* **155**, 75–82 (1969).
10. Clayton, D.D. *Astrophys. J.* **188**, 155–159 (1974).
11. Chan, K.W. & Lingfelter, R.E. *Astrophys. J.* **318**, L51–L55 (1987).
12. Gehrels, N., MacCallum, C.J. & Leventhal, M. *Astrophys. J.* **320**, L19–L22 (1987).

Donald D. Clayton is in the Department of Physics, University of Durham, DH1 3LE, UK, and is on leave from Rice University, Houston, Texas, USA.

Biochemistry

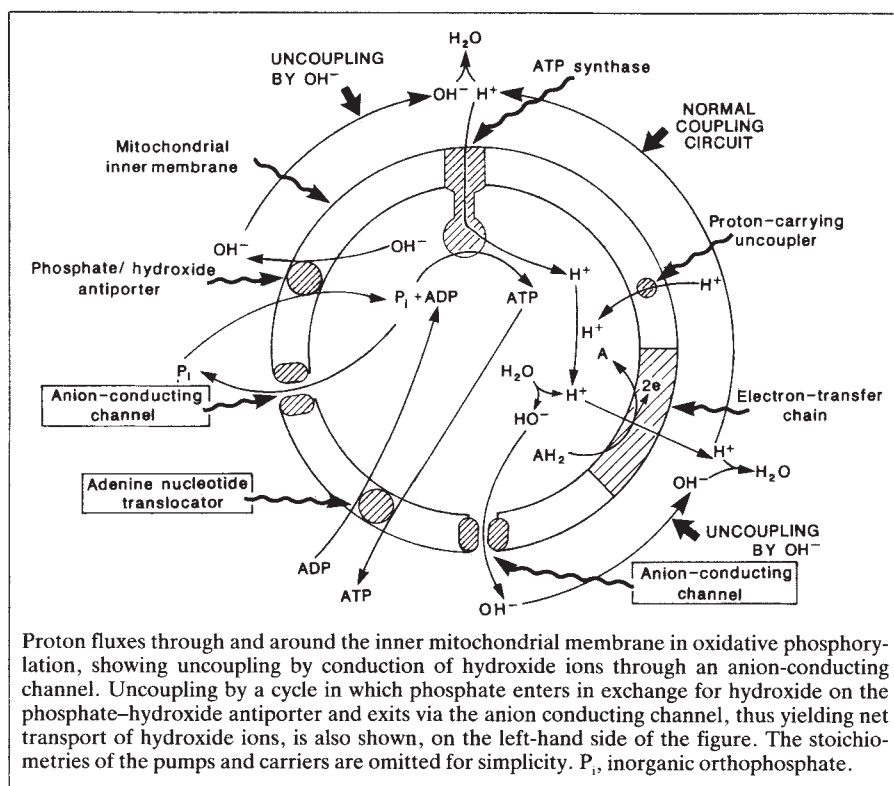
Holes in mitochondrial inner membranes

M.J. Selwyn

ONE of the requirements of Mitchell's chemiosmotic mechanism of oxidative phosphorylation, in which fluxes of protons couple oxidation reactions to ATP synthesis, is that the inner mitochondrial membrane is impermeable to protons. With aqueous phases on both sides of the membrane, hydroxide ions travelling in one direction are equivalent to protons travelling in the other, and a channel that conducts hydroxide ions will uncouple the system (see figure). It may therefore come as a shock, particularly to those who favour the delocalized version of chemiosmotic coupling, in which protons equilibrate with the bulk aqueous phases, to see the report by Sorgato, Keller and Stühmer on page 498 of this issue¹ of a direct demonstration of an anion-conducting channel in the mitochondrial inner

membrane. The authors treated mice with cuprizone to obtain giant mitochondria, removed the outer mitochondrial membrane and used the patch-clamping technique to seal a microelectrode onto the exposed inner membrane. This enabled them to measure the membrane potential and the ion-conducting properties of single channels in the membrane. This is a significant achievement, because the small size of mitochondria means that classical electrophysiological techniques using microelectrodes inserted into mitochondria give equivocal results. Their direct demonstration¹ of an inner membrane channel is not only intrinsically valuable but adds an important experimental technique at a time when there is growing interest in such channels.

The anion-conducting channel in the



inner membrane of mitochondria from brown adipose tissue is well characterized. When open, it uncouples the mitochondria and the consequent heat production is important in cold adaptation and in hibernating animals. The protein forming the channel, aptly named thermogenin or uncoupling protein, has been isolated and re-incorporated into artificial membranes. Similar definitive biochemical demonstration of anion-conducting channels in other mitochondria is lacking, but several groups have shown that mitochondria become permeable to anions (but not to non-electrolytes or zwitterions) under conditions of energization or when the external pH is alkaline. These results led to the hypothesis² of a pH-dependent anion-conducting pore or channel in the inner membrane of liver and heart

mitochondria. Apart from the lack of control by nucleotides, this channel resembles the channel in mitochondria from brown adipose tissue. Its role has remained obscure, although it was thought that it could act as a safety-valve if the proton-motive force rises too high, as it would open at high intra-mitochondrial pH and release internal hydroxide ions. The accumulating circumstantial evidence for this anion-conducting channel has recently been reviewed³ by Garlid and Beavis, who suggest other roles, including the restoration of mitochondrial volume after pathological swelling of the mitochondria.

A group in the Soviet Union finds⁴ that a heat-stable cytoplasmic factor stimulates the activity of the pH-dependent anion-conducting channel and these authors suggest a phosphate cycle, shown in the figure, which might be faster than direct hydroxide transport, particularly at neutral pH when the hydroxide concentration is very low. The amount of this factor is increased by treatment of rats with thyroid hormone⁵. This ties in with other recent reports of effects of thyroid hormones on mitochondrial function (see refs 6,7), and is interesting in that it suggests a molecular basis for the control of heat production by thyroid hormone and a physiological role for the anion-conducting channel in liver and heart mitochondria.

Another possible role for the channel is protection of mitochondria under conditions of oxygen stress — both anoxia (in ischaemia, for example) and hyperoxia. Proton motive force is retained during anoxia and the mitochondrial respiratory activity is suppressed after short-term

anoxia^{8,9}. The conductivity of the pH-dependent anion-conducting channel is about 20-fold greater after storage of the mitochondria in aerobic conditions than after storage in nitrogen¹⁰. *In vivo* this could function as a regulatory mechanism to minimize dissipation of the observed proton motive force under anoxic conditions^{8,9} and as a protection against potentially harmful high oxygen concentrations by uncoupling oxidative phosphorylation and thereby increasing oxygen use.

A different pore has recently been described¹¹. It is not specific to anions, allowing passage of sucrose, it is induced by calcium, particularly in the presence of inorganic orthophosphate, and its activity is also enhanced by oxygen stress, produced in this case by peroxides.

The channel described by Sorgato *et al.* in this issue¹ has features that differ from both the pH-dependent anion-specific channel and the calcium-induced pore, but the differences in experimental conditions and the use of the artificially induced giant mitochondria make it impossible to be certain that it is a different channel.

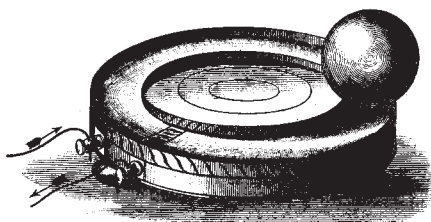
The situation is confused but exciting. The concept of the coupling membrane as an inert proton barrier now has to be discarded in favour of a membrane with channels that control the proton motive force. But, because the opening of the channels is controlled, their existence does not invalidate the concept of delocalized proton flows.

Thus, there seem to be three channels whose nature and separate existence require investigation. Evolutionary relationships between thermogenin, the adenine-nucleotide translocator and the phosphate-hydroxide antiporter have been proposed on the basis of sequence similarities^{12,13}, while the functional similarities of the pH-dependent anion-conducting channel and the thermogenin channel suggest they too are related. □

1. Sorgato, M.C., Keller, B.U. & Stühmer, W. *Nature* **330**, 498–500 (1987).
2. Selwyn, M.J., Dawson, A.P. & Fulton, D.V. *Biochem. Soc. Trans.* **7**, 216–219 (1979).
3. Garlid, K.D. & Beavis, A.D. *Biochim. biophys. Acta* **853**, 187–204 (1987).
4. Gainutdinov, M.Kh., Konov, V.V., Safarov, K.S. & Kasymov, A.K. *Biokhimiya* **51**, 387–394 (1986).
5. Gainutdinov, M.Kh., Mirmakhmudova, S.I., Mullagalieva, N.M. & Turakulov, Ya.Kh. *Byull. Eksp. Biol. Med.* No.3, 293–295 (1984).
6. Higui, T. & Rottenberg, H. *Chem. Pharm. Bull., Tokyo* **34**, 4331–4334 (1986).
7. Crespo-Armas, A. & Mowbray, J. *Biochem. J.* **241**, 657–661 (1987).
8. Andersson, B.S., Aw, T.Y. & Jones, D.P. *Am. J. Physiol.* **252**, C349–C355 (1987).
9. Aw, T.Y., Andersson, B.S. & Jones, D.P. *Am. J. Physiol.* **252**, C362–C368 (1987).
10. Selwyn, M.J., Comerford, J.G., Dawson, A.P. & Fulton, D.V. *Biochem. Soc. Trans.* **14**, 1043–44 (1986).
11. Crompton, M., Costi, A. & Hayat, L. *Biochem. J.* **245**, 915–918 (1987).
12. Aquila, H., Link, T.A. & Klingenberg, M. *FEBS Lett.* **212**, 1–9 (1987).
13. Runswick, M.J., Powell, S.J., Nyren, P. & Walker, J.E. *EMBO J.* **6**, 1367–1373 (1987).

M. J. Selwyn is in the School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK.

100 years ago MODERN VIEWS OF ELECTRICITY



Gore's circular railway. The light spherical metal ball revolves round the two concentric metal hoops or rails whenever it is made to convey a current between them in a vertical magnetic field.

From *Nature* **37**, 107; 1 December 1887.

Cosmology

Ever more distant quasars?

Peter Shaver

THE distances to far-off objects in the expanding Universe are measured by their large redshifts (z) — the fractional increases in the wavelength of the light they emit. There are now six quasars known with redshifts greater than 4; all have been found within the past year. Two of them, including the quasar with the highest-known redshift ($z = 4.43$), are reported on page 453 of this issue¹ by Warren *et al.* This rapid proliferation of high-redshift quasars is largely the result of improved search techniques, whose redshift limits are already being approached. It demonstrates that a redshift cutoff has not yet been found, and suggests that the space density of quasars may not decrease substantially towards these higher redshifts.

The space density of quasars above a given luminosity increases rapidly up to $z \approx 2$. The first quasar with $z \geq 2$ was only the ninth quasar known. Although the number of known quasars increased rapidly, the highest redshifts for several years were still not much above 2, giving rise to the notion of a redshift cutoff at $z \approx 2$ and speculation as to the possible cause. The quasar OQ172, found in 1973, has a redshift of 3.53, and held the record for almost a decade, during which time more than 1,500 quasars were found. This record was finally broken in 1982 by PKS 2000–330, which has a redshift of 3.78. Such high redshifts were, however, apparently rare, and it became widely believed that the space density of quasars peaked at $z \approx 2$, the 'quasar epoch' and fell off at higher redshift (light from distant objects must have been emitted a correspondingly long time ago).

Optical searches

All these redshift records were set by radio-loud quasars discovered through their radio emission; but most quasars are radio-quiet. Optical techniques are extremely efficient for finding quasars, but, until recently, have been biased against finding quasars with $z > 3$. One such technique is based on broadband colours: quasars have relatively blue continua, so 'UVX' quasar candidates are selected on the basis of an ultraviolet excess. The other technique is based on slitless spectroscopy: quasars have prominent emission lines, and 'objective prism' or 'grism' candidates are identified on the basis of both blue continua and characteristic emission features in low dispersion spectra. Both techniques are readily applied to photographic plates covering large areas, and digital scanning of these plates with computer analysis increases

the efficiency of these searches.

At higher redshifts the most prominent emission lines are shifted far to the red, and the ultraviolet excess disappears because of strong absorption by intervening hydrogen. Several recent quasar searches have therefore been made using both broadband colours and slitless spectroscopy at longer wavelengths, some using red-sensitive photographic plates^{1–3} and others scanning charge-coupled devices⁴; these make possible the detection of quasars out to redshifts of 4.5–5.0.

The promise of these new techniques has been realized with the recent burst of high-redshift discoveries (see table), and

Known quasars with $z > 4$;
 m_R , magnitude (brightness) in the red.

Quasar	z	m_R	Method	Ref.
0046–293	4.01	19	colour	2
0910+564	4.04	21	spectroscopy	4
0101–304	4.07	19.5	colour	1
0000–263	4.11	≈ 18	spectroscopy	*
2203+292	4.40	21	serendipity	†
0051–279	4.43	20	colour	1

* Hazard, C. *et al.*, in preparation.

† McCarthy, P. *et al.*, in preparation.

much of the redshift range over which they are sensitive has already been covered. These high-redshift quasars are relatively luminous, although not the most luminous known. They also appear to be surprisingly abundant, judging from the fact that three of them were found on just one $6^\circ \times 6^\circ$ plate, and the amazing serendipitous discovery by P. McCarthy and M. Dickinson (personal communication) of another whose spectrum was obtained by chance during a long-slit observation of a $z = 0.8$ radio galaxy. This is consistent with some recent analyses according to which the space density of bright quasars may still be increasing at high redshifts^{3,5}.

The observed space density of high-redshift quasars has a direct bearing on such issues as galaxy formation, obscuration by intervening matter, and gravitational lensing — in addition to the evolution of quasars themselves. The very existence of quasars of high redshift is awkward for some theories which predict relatively late formation of galaxies, such as those based on cold, dark matter. The spectra of the high-redshift quasars are not markedly different from those of others at lower redshifts, and the presence of similar heavy-element absorption lines due to intervening high-redshift matter indicates enrichment by still earlier generations of stars. The first galaxies

apparently formed well before the time corresponding to $z \approx 4$ when the Universe was less than 10–20 per cent of its present age.

Studies of faint galaxies and absorption systems in quasar spectra indicate that galaxies up to $z \approx 4$ –5 may cover almost the whole sky. If these contain dust they could obscure distant quasars⁶, and this could explain the fact that the space density of known quasars is insufficient to produce the high degree of ionization of the intergalactic medium at high redshifts. The known high-redshift quasars, however, do not appear to be significantly reddened by absorption⁷ (with the possible exception of Q0051–279 itself), and any actual redshift cutoff due to obscuration must evidently be in excess of $z \approx 4$ –5.

Gravitational lensing

The opposite effect, an apparent increase in the space density of luminous quasars, could conceivably be produced through magnification by gravitational lensing. Progress in understanding these issues will require the laborious assembly of statistically complete samples of quasars at these high redshifts, now possible with the new techniques.

What are the prospects for pursuing the quasar search to still higher redshifts? Further adaptation of the above methods, augmented by near-infrared photometry and narrowband imaging, could extend the accessible redshift range to $z \approx 6$. At $z = 7$, hydrogen Lyman- α lines would be redshifted to almost $1 \mu\text{m}$, and the entire optical band would probably be heavily obscured by absorption by hydrogen along the line of sight. Spectroscopic confirmation, as a result, would have to be made in the near infrared. Searches could in principle also be made in the infrared, although the relevant technology is still in its infancy.

Alternatively, one could concentrate on 'blank-field' X-ray or radio sources: that is, those with no optical counterparts to very faint limits. Virtually all quasars are X-ray sources, but those at high redshifts will be weak, and a sufficiently sensitive satellite has yet to be launched. Blank-field radio sources offer more immediate prospects, although the number of radio-loud quasars at high redshifts is probably very small. The recent successes at redshifts of 4–5 will undoubtedly stimulate such efforts. □

1. Warren, S.J., Hewett, P.C., Osmer, P.S. & Irwin, M.J. *Nature* **330**, 453–455 (1987).
2. Warren, S.J. *et al.* *Nature* **325**, 131–133 (1987).
3. Hazard, C., McMahon, R.G. & Sargent, W.L.W. *Nature* **322**, 38–40 (1986).
4. Schmidt, M., Schneider, D.P. & Gunn, J.E. *Astrophys. J.* **321**, L7–L10 (1987).
5. Koo, D.C. & Kron, R.G. *Astrophys. J.* (in the press).
6. Ostriker, J.P. & Heisler, J. *Astrophys. J.* **278**, 1–10 (1984).
7. Wright, E.L. & Malkan, M.A. *Bull. Am. astr. Soc.* **19**, 699 (1987).

Peter Shaver is at the European Southern Observatory, Karl-Schwarzschild-Strasse 2, D-8046 Garching bei München, FRG.

Detection of stratospheric nitrogen species

SIR—When Toon *et al.*¹ claimed a definitive detection of N_2O_5 in the stratosphere from atmospheric trace molecule spectroscopy (ATMOS) data, they showed only one absorption band that could be attributed to N_2O_5 . They claimed another band near $1,700\text{ cm}^{-1}$ was evident, but this claim might be thought dubious because there are strong overlapping bands of HNO_3 at $1,700\text{ cm}^{-1}$, which differ between Northern Hemisphere (sunset) and Southern Hemisphere (sunrise) spectra, and the authors did not show this other band. My own tentative measurement of N_2O_5 (ref. 2) has been described by Camy-Peyret and others as a measurement of “an N_2O_5 -like substance.” So far, the ATMOS detection could rightly be claimed to be one of “an even more N_2O_5 -like substance”, but it is not an unambiguous detection as the authors claim. The evidence for it being N_2O_5 is very good, but many constituents have P- and R-branch features like the observed absorption, and although none is known at the observed wavelength, chemists are not infallible at predicting which trace constituents will or will not exist.

Most spectroscopists would agree that a definitive identification can only be claimed when two isolated or well-resolved lines or bands are observed simultaneously. This point was clearly stated by Rinsland *et al.*³, who analysed balloon-borne spectra to search for absorption by $ClONO_2$ —they declined to claim a detection of $ClONO_2$ because they observed two bands, but on different balloon flights. This was left for the ATMOS team⁴ who analysed four bands of $ClONO_2$ and published two. The ATMOS team has clearly taken the point because its tentative detection of HO_2NO_2 (ref. 5), which consisted of one band only, was entitled “Evidence for the presence of . . .”.

Over a year has elapsed, and Toon *et al.* have still not published additional spectra to establish their claim. Yet they continue to assert that they have definitively measured N_2O_5 , most recently in a paper submitted to the *Journal of Geophysical Research*⁶. Furthermore, administrators

have accepted their assertions: in their 1986 report to the United Nations Environment Programme (UNEP), the Coordinating Committee on the Ozone Layer claimed that “the presence of the temporary reservoir species $ClONO_2$, N_2O_5 and HO_2NO_2 in the stratosphere has been definitively established by recent observations”. This is false on two counts.

I am as keen as anyone to establish that the broad absorption band observed in the atmosphere at $1,240\text{ cm}^{-1}$ arises from N_2O_5 because I can then continue to measure it with my radiometer and label the measurements ‘ N_2O_5 ’ without the ‘-like substance’. But the authors do us all a disservice by claiming a positive identification without the proper evidence.

H.K. ROSCOE

Department of Atmospheric Physics,
Clarendon Laboratory,
Parks Road, Oxford OX1 3PU, UK

1. Toon, G.C., Farmer, C.B. & Norton, R.H. *Nature* **319**, 570–571 (1986).
2. Roscoe, H.K. *Geophys. Res. Lett.* **9**, 901 (1982).
3. Rinsland, C.P. *et al. J. geophys. Res.* **90**, 7931–7943 (1985).
4. Zander, R., Rinsland, C.P., Farmer, C.B., Brown, L.R. & Norton, R.H. *Geophys. Res. Lett.* **13**, 757–760 (1986).
5. Rinsland, C.P. *et al. Geophys. Res. Lett.* **13**, 761–764 (1986).
6. Russell, J.M. III *et al. J. geophys. Res.* (submitted).

Toon replies—When we submitted our paper¹ we felt that an additional figure showing the $1,725\text{ cm}^{-1}$ band of N_2O_5 would add little to the paper because of its similarity to the existing figure of the $1,250\text{ cm}^{-1}$ band and because the identification was less clear-cut. I agree with Dr Roscoe that there are strong HNO_3 absorptions in the same spectral region as the $1,725\text{ cm}^{-1}$ N_2O_5 band, but the fact that the N_2O_5 layer peaks at a higher altitude than that of HNO_3 allows the N_2O_5 band to be observed without excessive interference over a narrow range of tangent heights (30–35 km). Figure 1 shows four spectra taken in this tangent-height range, two at sunrise and two at sunset. Because N_2O_5 is almost completely photolysed during the daytime, the sunrise spectra should contain its spectral signature and the sunset spectra should not. Figure 2 shows, superimposed upon the 30.6-km sunrise spectrum, separate calculations of the spectral signatures of N_2O_5 , HNO_3 and O_3 in the same observation geometry. All of these features can be identified in the spectrum; the additional strong, randomly posi-

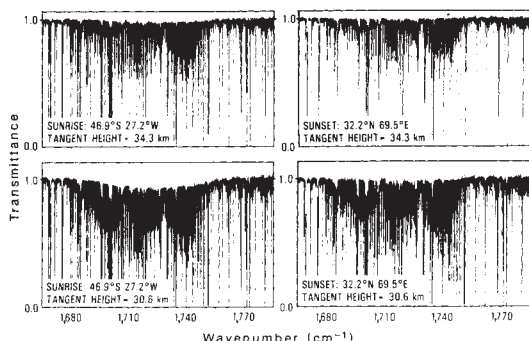


Fig. 1 Four infrared atmospheric transmittance spectra measured by the ATMOS sensor from 262-km altitude on 30 April 1985 (sunset) and 1 May 1985 (sunrise).

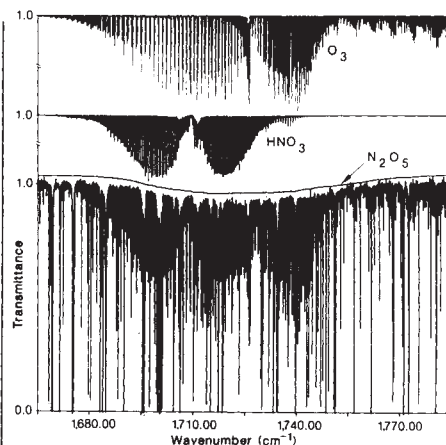


Fig. 2 A superposition of the 30.6-km sunrise spectrum of Fig. 1 and calculated spectra of N_2O_5 , HNO_3 and O_3 offset for clarity. Absorption of all three molecules can be identified in the sunrise spectrum beneath the confusion of strong water-vapour lines.

tioned spectral lines result from water vapour. Although the sunrise–sunset difference of the measured spectra does not conform exactly to the N_2O_5 signature, the wings of the latter being too large, only the most sceptical critic could fail to attribute this difference to N_2O_5 . The calculated spectrum of the $1,725\text{ cm}^{-1}$ N_2O_5 band was based on the room temperature laboratory measurements of Murcray *et al.*², with all spectral lines assumed to have the same ground state energy. Had we given these spectral lines a more realistic temperature dependence, by increasing their ground state energies away from the centre of the band (as Toon *et al.*¹ did for the $1,250\text{ cm}^{-1}$ band) the calculated wings of the $1,725\text{ cm}^{-1}$ band would have been weaker at stratospheric temperatures and their agreement with the measurement perhaps even better. This emphasizes the need for accurate low-temperature, laboratory measurements of both of these N_2O_5 bands so that the untapped information about N_2O_5 contained in the ATMOS spectra can be accurately reduced to volume-mixing-ratio profiles.

G. TOON

Jet Propulsion Laboratory,
California Institute of Technology,
4800 Oak Grove Drive,
Pasadena, California 91109, USA

1. Toon, G.C., Farmer, C.B. & Norton, R.H. *Nature* **319**, 570–571 (1986).
2. Murcray, D.J., Murcray, F.J., Goldman, A., Bonano, F.S. & Blatherwick, R.D. *High Resolution Infrared Laboratory Spectra* (Dept. Physics, Univ. Denver, 1984).

Incubation period for AIDS

SIR—Comments on my article, “The sombre view of AIDS”¹ have been uniformly critical. I would include in this a recent article by Medley *et al.*² on the AIDS (acquired immune deficiency syndrome) incubation period. Nearly all the criticisms focus on the estimation of the

AIDS density function — the probability of developing AIDS in time following infection — rather than on the estimation of AIDS infections from cases.

A major problem with the use of transfusion data for estimating the density function is the understatement of recent diagnoses, the effect of which is shown in Table 1, which presents adult transfusion cases infected in 1979–81, and reported to the Centers for Disease Control (CDC) by August 1985, April 1986 and April 1987. If one considers only the 1985 figures, one might conclude that the centre of the underlying distribution is around 4–5 yr; just this conclusion was reached by Lui *et al.*³, Iverson and Engen⁴, Beal⁵, and Costagliola and Downs⁶.

The figures for the later years show the dangers of this, for in the 1986 report the centre of the distribution seems to be about 5–6 yr, while by 1987 it seems to be between 6 and 7 yr. It is not surprising that Lui *et al.*⁷ are having second thoughts about their 4.5-yr mean.

Medley *et al.*² estimate an 8.23-yr mean by fitting a Weibull distribution to adult (5–59 yr at transfusion) cases. I have pointed out⁸ that the implications of this model are inconsistent with our current knowledge of the epidemiology of AIDS.

Another distribution, the gamma, was also considered. Medley and others state, “without independent information it is impossible to distinguish between the distributions in the current analysis”. It is misleading, therefore, to imply that the two distributions generate broadly consistent results, and disingenuous to fail to point out that their adult case fit with the gamma suggests a mean incubation period of 27.21 yr. For adult cases, the geometric mean of their gamma and Weibull means is 14.96 yr, providing support for the idea of a 15-yr mean. Unfortunately, no attempt was made to use the normal or log-normal distributions.

Barton⁹ has suggested a simple ‘actuarial’ test for normality by comparing the AIDS case ratio in the second and first years following transfusion with the third and second year ratio, and so on. I have reworked this test with 1987 data. The sequence I obtained was 4.19, 1.91, 1.16, 1.11, 2.00, 1.62, 0.00, and the expected

Table 2 US AIDS infection estimates

Yr	Number of cases	Yr	a	Infections b	c
1979	13	1978–89	10,744	2,093	2,708
1980	50	1979–80	22,588	— 103	— 923
1981	265	1980–81	148,212	17,306	32,808
1982	1,012	1981–82	461,128	41,499	39,650
1983	2,805	1982–83	895,281	1,445	102,698
1984	5,627	1983–84	775,017	–128,099	– 47,587
1985	9,757	1984–85	419,224	–224,017	151,960
Totals	19,529		2,732,194	–289,876	281,314

CDC data, omitting 12 pre-1981 cases. a, Normal model with $\mu = 15$ yr, $\sigma = 5$ yr; b, Normal model with $\mu = 5$ yr, $\sigma = 1.6$ yr; c, Weibull model² with $\mu = 8.2$ yr. Negative values set at zero during inversion.

sequence is 1.74, 1.68, 1.61, 1.55, 1.49, 1.43, 1.38. Bearing in mind that the first ratio is anomalous, because few AIDS cases occur in the first few months after infection, and that the final ratio is based on almost no cases, there seems no reason to reject the 15-yr mean normal model that I suggested, and Barton’s suggestion of a 5-yr mean is unlikely.

Several writers have pointed to my omission of maximum likelihood, significance tests and confidence intervals. My aim was to show that the selected model was consistent with the data, not that it was an optimal description of them. Given the quality of the transfusion data, and our uncertainty about underlying statistical processes, the use of such methods is premature, and suggests a precision that does not exist.

Absolute rather than squared deviations were used as a simple method of matching model to data. I did investigate a range of standard deviations but, as can happen, the table was edited out.

In Table 2 there is a comparison of the estimated yearly growth of infections with human immunodeficiency virus (HIV) suggested by the 15- and 5-yr mean normal models and Medley’s 8.23-yr mean Weibull models. Neither the 5-yr mean normal, nor the 8.23-yr mean Weibull, are capable of generating sensible estimates for the time series of HIV infections, whereas the 15-yr normal produces estimates that are reasonable, even if some think them on the high side. Both the 5-yr mean normal and the 8.23-yr mean Weibull suggest that new infections

were negative in some years, an obvious impossibility.

The 15-yr mean normal model shows a peak in new infections in 1982–83 and a subsequent decline, a pattern that is broadly consistent with the few new infections in London among homosexuals from 1984¹⁰. Such a pattern can be explained using a three-stage infectiousness model, with a short initial period of infectiousness, followed by non-infectious dormancy, and then a longer final period of infectiousness at the AIDS stage. The 8.23-yr mean Weibull, however, peaks in 1984–85, after a period of negative infections in 1983–84, and another smaller peak in 1982–83. This cannot be easily explained.

Costagliola and Downs⁶ imply that idea that nearly all those infected with the HIV will develop AIDS or terminal HIV disease unless adequate treatment is developed, is “startling”, and quote with approval an estimate by Iversen that 27% of seropositives can be expected to develop AIDS. When my paper was first written in early 1986 it was common to quote 10%; now 30% is popular. Cohort studies in the United States suggest that prognosis is getting worse for those infected, although no evidence is emerging of a subgroup of HIV-infected cases who do not go on to develop AIDS.

Medley *et al.* refer to “infective transfusions” when they really mean transfusions that will result in AIDS, barring premature death from something else. This shows the change of thinking which is occurring.

MALCOLM REES

John Hunter Clinic,
St Stephen’s Hospital,
Fulham Road,
London SW10 9TH, UK

1. Rees, M. *Nature* **326**, 343–345 (1987).
2. Medley, G.F., Anderson, R.M., Cox, D.R. & Billard, L. *Nature* **328**, 719–721 (1987).
3. Lui, K.-J. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 3051–3055 (1986).
4. Iversen, O.-J. & Engen, S. *Epid. Comm. Hlth* **41**, 55–58 (1986).
5. Beal, S. *Nature* **328**, 734 (1987).
6. Costagliola, D. & Downs, A.M. *Nature* **328**, 582 (1987).
7. Lui, K.-J., Peterman, T.A. & Lawrence, D.N. *Nature* **329**, 207 (1987).
8. Rees M. *Lancet* **ii**, 98–99 (1987).
9. Barton, D.E. *Nature* **326**, 734 (1987).
10. Carne, C.A. *et al. Lancet* **i**, 656–658 (1987).

Table 1 Transfusion cases as reported by CDC

Year of infection	1979			1980			1981		
	1985	1986	1987	1985	1986	1987	1985	1986	1987
Incubation period to AIDS (yr)									
1	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	6	6	6
3	1	1	1	8	8	7	8	12	13
4	1	3	3	11	11	10	18	20	21
5	6	4	4	4	8	7		13	27
6	4	7	6		6	14			16
7		5	10			7			
8			5						

Do-it-yourself science

Alan Glynn

Who Goes First? The Story of Self-Experimentation in Medicine.

By Lawrence K. Altman. Random House: 1987. Pp. 430. \$22.50.

As Dr Johnson said in another context, "It is not done well; but you are surprised to find it done at all". One gets the same feeling, at any rate to start with, on reading some of the experiments that people have carried out on themselves. Altman's contribution is to show us not just that a few famous experiments have been of great scientific value but how extensive the whole process of self-experimentation is and, indeed, how necessary from an ethical point of view. If you are not prepared to experiment on yourself, how can you ask others?

Like all good books, this one can be read at several levels. At the most superficial it provides a good read for almost everyone. There is drama and mystery. What, for example, would Poirot have made of the case of the Canadian physiologist found alone and unconscious in the laboratory one Saturday evening? There are plenty of heroes and eccentrics, often combined. Castle who sorted out the intrinsic factor story in pernicious anaemia, was also a handyman and it was sometimes easier to consult one of the most illustrious professors at Harvard by asking him to mend a tap than to see a patient. On the other hand there are no real villains and no sex except for a few worried wives in the background. Indeed there are very few women in the book at all. Altman suggests this is a reflection of the relatively small number of women in medicine or medical research until recently. He is probably right (though there is also the possibility that women have more common sense than men). It is interesting but perhaps not significant that the woman who gets the most mention, Dr Elsie Widdowson, was a nutritionist. Mrs Goldberger insisted on sharing her husband's self-experimentation into the causes of pellagra in order to represent women.

Nutrition, drugs, physiology and infection are the main fields covered, with the latter by far the largest. The risks and discomforts of each differ. Santorio of Padua spent 30 years of his life with bed and table on a steelyard to measure the weight changes due to what we would now call metabolism; tedious but hardly

dangerous. Others were not so lucky. In 1769 William Stark tried to see if simple diets were healthy, and died of scurvy, though James Lind had described the remedy 16 years before.

Serendipity played its part. While investigating the properties of cocaine, Freud noticed a numbness of the tongue. He mentioned it to his friend, the ophthalmologist Koller, who developed

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Taking a chance — earlier this year Daniel Zagury caused a sensation when he injected himself with a new AIDS vaccine.

the use of cocaine as a local anaesthetic. Likewise, one day in Egypt Arthur Loos was surprised to find himself infected with hookworm, then thought to be spread only by ingestion. However, he remembered the burning sensation when a drop of a larval culture had fallen on his hand and went on to demonstrate experimentally the penetration of his intact skin by larvae of *Ankylostomum duodenale*.

There must be something about major advances. Loos's finding was not accepted for some time, because J.S. Haldane and Boycott could not at first confirm the results. This recalls the delay in the mid-nineteenth century in accepting Villemin's classical demonstration of the transmissibility of tuberculosis because an English Commission of distinguished scientists could not confirm his work. Or rather they did, but they managed also to infect one of the control guinea pigs, so coming to the wrong conclusions.

Pettenkofer, though fortunate in only getting a mild diarrhoea after drinking a beakerful of cholera vibrios, was unlucky, because he did not recognize his illness and so was confirmed in his view that the micro-organism alone is insufficient. There is some truth in this, but unfortunately he went further and assumed he had demonstrated the need for a theoretical soil factor, in which he passionately believed. Of course he was anxious to prove a point, but such an approach is not confined to self-experimenters. Complete scientific detachment is rare, though the courage to test scientific theories is clearly not.

Dr Altman traces back his interest in self-experimentation nearly 30 years and his qualities are reflected in several

aspects of his book. As a physician with some experimental tendencies he gives an accurate picture of the medical and scientific background to the experiments he describes, so allowing both himself and us to judge their proper worth. As a scholar he presents the detailed results of his research in a field where facts are difficult to come by. As a full-time journalist his writing is relaxed and interesting to read, with much human detail which hovers between verification of history from living sources and the journalistic interview. Just occasionally the latter takes over.

There are some interesting questions. What is the true story of John Hunter's auto-inoculation? Why did Walter Reed go to Washington while his colleagues were being bitten by yellow-fever-bearing mosquitoes? What were the motives of all these people? At a time when medicine is too often dismissed as technology, and anti-science — based as always on ignorance and superstition — abounds, it is refreshing to read of men and women whose driving force was intellectual curiosity applied to problems of medicine no matter what the cost. To J.B.S. Haldane, the intellectual giant, it would have been an ideal way of dying. Pettenkofer and Metchnikoff among others, would have agreed. Dr Altman's book has itself gone first in exploring his subject. The questions he raises may be answered by some future student from the archives of the Institutions Review Boards and ethics committees, themselves inspired by the shadow of Nuremburg. □

Alan Glynn is Director of the Central Public Health Laboratory, Colindale, London NW9 5HT, UK, and a visiting Professor at the London School of Hygiene and Tropical Medicine, University of London.

Building and design

Janet M. Thornton

Protein Engineering. Edited by Dale L. Oxender and D. Fred Fox. Alan R. Liss: 1987. Pp.365. \$36, £25.75.

SUCCESSFUL protein engineering requires expertise in two distinct disciplines, molecular genetics and structural biophysics. A protein engineer should, at minimum, understand the basic principles of each of them. This text attempts to provide an introduction to both subjects, although it includes substantially more biophysics than genetics.

The book was conceived at a Genex-UCLA Symposium in 1985, with the aim of providing a supplement to introductory biochemical textbooks. It includes contributions from 30 laboratories, grouped into four sections. The first two describe basic methodology for structure determination, gene manipulation and protein biochemical analysis, and are followed by a section on "Energetics and Protein Design". The last section, "Purposely Modified Proteins and Their Properties", includes results of experiments on particular proteins. Each contribution is short (on average ten pages) and, to provide cohesion, overviews of each section written by eminent scientists are included. This ploy is only partially successful and the text resembles a collection of short stories rather than a novel.

The quality and level of the chapters is variable, although most are good and a few are outstanding. The methodology

contributions are necessarily introductory; for example, in nine pages it is possible to describe X-ray diffraction only at a rudimentary level. There is also some repetition, especially with regard to basic structural principles and our lack of understanding of the relationship between sequence, structure and function.

The multi-author approach comes into its own when the contributors discuss results from specific protein and peptide systems. This book brings together accounts of some of the first successful experiments on modifying and designing proteins. Many of these contributions are informally and fluently written and capture some of the authors' doubts and excitement in designing novel molecules. They are complementary to the standard descriptions to be found elsewhere and provide interesting reviews for budding protein engineers and lecturers in this field. For example we are treated to an insight into the approach and problems of designing completely novel molecules, the design of 'minimal' amphiphilic peptides which retain biological activity, the detailed analyses of the contribution of specific residues to catalytic activity and attempts to increase protein stability.

This book provides only a limited introduction to the detailed methodology involved in protein engineering. However, the non-specialist will find in it a delightful assortment of some of the first and most interesting experiments performed by protein engineers. □

Janet M. Thornton is a Lecturer in the Department of Crystallography, Birkbeck College, University of London, Malet Street, London WC1E 7HX, UK.

New light on signals and connections

David Attwell

The Retina: An Approachable Part of the Brain. By John E. Dowling. Harvard University Press: 1987. Pp.271. \$37.50, £29.95.

THE retina has two functions: it converts light into an electrical signal in rods and cones and then it processes this signal, with four layers of neurons, before despatching it along the optic nerve to the brain. Compared to other parts of the central nervous system, the retina ought to be easy to understand. Its input signal, the visual image entering the eye, is known, and its output at the optic nerve can be measured relatively simply: the only problem is to work out what goes on in between, and why.

The Retina reviews our current understanding of 'what goes on in between', to which John Dowling and his collaborators have made major contributions over the past 25 years. The book includes an impressive and amply illustrated summary of the anatomy of retinal neurons and the synapses between them, a description of the electrical responses of the retinal cells, and chapters on synaptic pharmacology, visual adaptation and phototransduction, and the electroretinogram and glial cells. The last book of this scope devoted to the retina, R.W. Rodieck's excellent *The Vertebrate Retina*, published by W.H. Freeman in 1973, is now rather dated, so Dowling's book is sure to find a place in libraries and on researchers' shelves.

Nevertheless, I found two aspects of the book — the overall structure and the weight given to different topics — rather unsatisfying. For the intended readers (presumably graduate students), a more systematic exposition of the subject matter would have been helpful. Inadequate linking of different sections of the book, and a tendency to present retinal anatomy in great detail without any functional context, often make it difficult to find a coherent account of one topic. For example, although the shape of horizontal cells is described on page 22, and their output synapses on page 60, it is only on page 110 that the physiological significance of this anatomy and synaptic connectivity is revealed. Ganglion cell light responses are presented in two unrelated sections, which describe extracellular and intracellular recordings. Similarly, although the reader is told briefly, after 80 pages, that the photoreceptors are hyperpolarized by light, an account of how this happens is deferred to the penultimate chapter.

In the preface, Dowling states that he



Steel sidewinder — the twisted railroad tracks at Lázaro Cárdenas on Mexico's Pacific coast are a vivid scar left by the earthquake of September 1985, which killed over 9,000 people. The picture is taken from *The National Geographic Society — 100 Years of Adventure and Discovery* by C.D.B. Bryan, a beautifully presented collection of photographs from the magazine's archives with accompanying commentary. The book is published in the United States by Abrams and in Britain by Phaidon.

has emphasized those studies which he thinks have given the most useful insight into retinal function. This sometimes leads to a rather unusual weight being given to different topics. Thus, on phototransduction, the visual pigment and its reactions following light absorption are described in detail over five pages, while the subsequent chain of reactions involving cyclic GMP (whose recent unravelling has been a major success in retinal research) are described cursorily in less than two; the recently disproved calcium hypothesis of phototransduction is not mentioned. Similarly, Enroth-Cugell and Robson's important experiment distinguishing X and Y ganglion cells is only thought to be worth describing in a footnote, while ten pages are devoted to Dowling's studies of the effects of dopamine on horizontal cells. Quantitative analyses of retinal function, such as Lamb and Simon's work on the electrical

coupling of retinal cells, are largely neglected.

Finally, the descriptions in this book of what occurs in the retina could have been accompanied by more analysis of *why* it occurs. Readers who want to know why the retina segregates visual information into ON and OFF channels, why there are sustained and transient ganglion cells, or why bipolar and ganglion cells have receptive fields with an excitatory centre and an inhibitory surround, will not find such issues discussed in any detail here, despite their relevance, for example, to Marr's theories of early visual processing.

This is a useful book. But it would have better served the graduate student entering the field if it had followed the philosophy of its subtitle, and given a broader view of retinal function. □

David Attwell is in the Department of Physiology, University College London, Gower Street, London WC1E 6BT, UK.

Reactions to risk

John Dunster

The Politics of Anxiety: Sellafield's Cancer-Link Controversy. By S.M. Macgill. *Pion, 207 Brondesbury Park, London NW2 5JN; 1987. Pp.198. Pbk £14.95.*

MOST of those involved in risk estimation and its practical application now recognize that the estimate of a risk and the fear of a risk are not closely related. The fear, or 'perception' in the jargon of the trade, is no longer seen as irrational or the result of ignorance; it is recognized as real, perhaps more real than the risk creating it. We appreciate the need to assess these fears and to understand how they arise. Creating them and exacerbating them is easy enough. We would like to know how to assuage them when they seem to be excessive. Even more, we would like to know how best to incorporate them into the process of managing and regulating risks. This book helps with the understanding but offers little help in its application.

On the whole, the book is carefully and clearly written, with few lapses into jargon, but the reader's patience may be strained throughout Chapter 1 by an excessive use of parentheses. The later chapters are mercifully less encumbered.

To me, the most fascinating parts of the book are those summarizing the reactions of the West Cumbrian public to the presence of the nuclear plant at Sellafield, to the Yorkshire Television documentary, *The Nuclear Laundry*, and to the Black Report of 1984 which was the result of an inquiry into the incidence of leukaemia around Sellafield. This type of information is extremely difficult to summarize and it says much for the author's clarity of

thought and presentation that these chapters are compelling reading. They seem to me, a layman in this area, to be examples of social science at its best.

The later chapters on risk assessment and on the logic and science of the Black Report impressed me less. It may be that the author has chosen to restrict her source material to the Black Report, but I suspect that a basic lack of expertise is more to blame. There is really no excuse for failing to recognize that the radiobiological approach offered to Black led to a forecast 400 times smaller than the observed excess rather than 40 times. The latter figure comes from a somewhat implausible modification made in the Black Report attributing *all* childhood leukaemia everywhere to radiation. The point is brought out clearly in Para 4.50 of the Report. There is some evidence of bias in this part of the book, probably stemming from an excessive dependence on the reports of Black and COMARE (Committee on the Medical Aspects of Radiation in the Environment).

Here, too, there are some weaknesses of logic. The author is puzzled that Black should think that a larger factor between the calculated and the observed number of leukaemias would decrease the likelihood that the releases from Sellafield have caused the leukaemias. She believes that a larger factor would strengthen the *a priori* case against Sellafield. She fails to distinguish between the existence of the works at Sellafield and the possible causal mechanism. She also appears to think that an excess that is statistically significant must necessarily have a cause. In fact, the excess at Seascale is unlikely to be due to blind chance, but the possibility cannot be excluded until a convincing explanation can be found. Either the author deals with physical and biological science less well

The ANS Journals

**YOUR
NUMBER ONE
SOURCE FOR CURRENT
INFORMATION ABOUT
NUCLEAR
ENERGY**

- ☐ NUCLEAR NEWS
- ☐ NUCLEAR SCIENCE & ENGINEERING
- ☐ NUCLEAR TECHNOLOGY
- ☐ TRANSACTIONS
- ☐ ANS NEWS
- ☐ REMOTE SYSTEMS TECHNOLOGY
- ☐ FUSION TECHNOLOGY

FOR OUR CATALOG
CALL OR WRITE:

AMERICAN NUCLEAR SOCIETY
555 N. Kensington Avenue
La Grange Park, IL 60525 USA
312-352-6611

#

Reader Service No.16

than with social science, or I am a harsher judge in the areas I know best.

The final chapter, "Summary and Outlook", is disappointing. The author sees her aim as "to reemphasise and so to bring more fully to the fore a number of underacknowledged, ignored, or newly emerging factors, with the intention of encouraging greater awareness of the true character of the context for future public policy for dealing with risk controversies". None of the factors dealt with in the book justify those adjectives, although it was worth re-stating them lest they be forgotten. But that is not the issue that makes this book worthwhile. Its value lies in showing that preconceived notions of what people think are suspect, and that the sources of information available to the general public are inadequate. It is not necessarily true that better and more complete information will lead to less unjustified fear and to less complacency, but without better information in better forms there is no way forward. Descriptive sociology, like descriptive biology, may be of limited use and is easy to criticize, but it is a vital starting point. This book is necessary reading. It is not, and could not be, sufficient. □

John Dunster, 52 Thames Street, St Ebbs, Oxford OX1 1SU, UK, recently retired as Director of the National Radiological Protection Board.

Supercontinental supposition

M.W. McElhinny

Palaeomagnetism and the Continental Crust. By J.D.A. Piper. *Open University Press, Milton Keynes/Halsted Press, New York*:1987. Pp.434. £35, \$59.95.

It is nearly 15 years since the appearance of the last major text on palaeomagnetism in which the available data were used to review past geometries and movements of the crust through geological time. There has been a great expansion of data in these years, and a book reviewing and updating the situation would generally be considered very timely. It was therefore with considerable interest that I read this new text of 374 pages plus 45 pages of references.

The first six chapters (150 pages) are an introduction to palaeomagnetism as a subject, and include discussion of the basic principles of rock and mineral magnetism, magnetic minerals, the theory of rock magnetism and a particularly good chapter on magnetization in rocks. The method of acquisition of magnetization by different rock types is presented in a way that should be helpful for those wishing to learn something of the background to an area that is generally not well understood by the non-specialist.

The following chapter on field and laboratory methods is less satisfactory. In particular, in the discussion of demagnetization techniques no attempt is made to relate theory and experiment in any substantive way; current procedures are complex, and the uninitiated will not get a feel for the intricacies of the analysis of demagnetization diagrams, a matter of great importance when results have to be interpreted. Things do not improve in a subsequent chapter when the identification and separation of magnetic components is considered in more detail.

In the following five chapters global palaeomagnetic data are analysed from Archaean to Cenozoic times, and the geological implications are discussed. Unfortunately (and in his introduction the author explicitly makes no apology for this) the synthesis is based entirely on the supposition that a single supercontinent existed throughout most of Precambrian time. All the analyses are based on this model and the resulting diagrams are extremely difficult to follow; for example, it is not easy to isolate in one's mind the data from any one continent as a set. There are also a large number of errors in the spelling of formation names, and altogether following these diagrams requires what I can only describe as intestinal fortitude.

Only a small minority of palaeomagnetists agree with the supercontinent model, and I personally do not believe the data support it in the way that Piper claims. The main criticism is that the analysis of Precambrian pole paths is still the same as that used a decade or more ago; in my view this is not now an appropriate method or approach to the subject. It would have been interesting if the author had produced some computer simulations of the effect on Precambrian pole paths of applying the various crustal models he puts forward in Chapter 6. It seems obvious that the analysis, if applied to Phanerozoic data, would have missed the opening of the Atlantic Ocean! Yet we are

continually led to believe the data support the supercontinent model and in many instances the contention becomes hard to sustain.

Piper has clearly done an enormous amount of work in putting the text and analysis together. But he presents a personal view, one that is not at all widely accepted by his peers and one which demands critical assessment on the part of the reader. Unfortunately such an assessment will not be at all easy for the non-specialist in palaeomagnetism. □

M.W. McElhinny is Chief of the Division of Geophysics, Bureau of Mineral Resources, Geology and Geophysics, GPO Box 378, Canberra 2601, Australia.

Out of sequence

W.J. Brammar

The Gene: Its Structure, Function and Evolution. By Lawrence S. Dillon. *Plenum*: 1987. Pp. 896. \$95 (North America); \$114 (elsewhere).

For obvious reasons, much of the exciting progress in modern biology has centred on 'the gene'. New insights into gene structure and function have been frequent and often surprising in the past decade. The potential for direct application of molecular genetics in biotechnology and its importance for the understanding of human disease, including cancer, has stimulated activity in the field and promises to provide patent lawyers with a steady source of income.

In this major new treatise, Lawrence Dillon has aimed to help our understanding "by thoroughly analyzing the genes and their surrounding sequences from a broad spectrum of organisms". His approach boils down to comparative analysis of nucleotide sequence data, by which he believes that "it is *inevitable* that new general characteristics and inter-relationships among genes . . . should be disclosed" (my italics).

Well, there's nothing wrong with being optimistic at the outset of a long and arduous journey; but the traveller had better be prepared for difficulties on the way and disappointment on arrival. Here many of the difficulties are unfortunately imposed by the author himself. Some authors are easy to read because they write well. William Hayes's early classic *The Genetics of Bacteria and Their Viruses* and Mark Ptashne's recent monograph *A Genetic Switch: Gene Control and Phage λ* provide contrasting examples of authoritative scientific books that were a pleasure to read.

Dillon manages the opposite effect, using and abusing words so profligately as to obscure the meaning of the text. He

introduces new uses for long-accepted terms: thus nucleotide pairs in the genetic sequence become "sites", while conventional sites, such as operator sequences or other *cis*-specific elements, are frequently referred to as genes. Structural genes, their transcripts and their encoded proteins are frequently transposed by careless and confusing use of the language.

All the known types of genes are chronicled and catalogued, with many of their sequences presented in lengthy tables. Dillon has an obsession with molecular phylogeny, going to extremes to try to fit gene sequences into some arcane classification. The subdivisions are often based more on the protein product than the gene itself, and brash new terms are coined in attempts to give them substance. We are introduced to "diplomorph" and "cryptomorph" genes, differentiated by the ways in which the gene products are post-translationally processed. The repetitive *Alu* elements, elevated to the status of genes, are subdivided into "monalus", "diplalus", "triplalus" and "analus" based on the number of discernible sequence-repetitions.

Dillon is not averse to admonishing the molecular biologist, often wrongly, for missing some seemingly important feature of sequence information. Unfortunately the coverage of the literature does not extend beyond 1985, so that many of the questions he poses have now been answered or supplanted. And although many of the sequences considered are involved in controlling gene expression, Dillon has chosen to give the topic of regulation only the most cursory treatment.

Nature's entreaty to reviewers that "Although book reviews should be informative, they should not be dull" would be equally sound advice to authors of scientific texts. □

W.J. Brammar is a Professor in the Department of Biochemistry, University of Leicester, Leicester LE1 7RH, UK, and Director of the University/ICI Joint Laboratory for Molecular Genetics.

Newtonianism, reductionism and the art of congressional testimony

Earlier this year, at the University of Cambridge, Steven Weinberg gave a talk at the Tercentenary Celebration of Newton's Principia. This is what he said.

MY TALK this afternoon will be about the philosophy of science, rather than about science itself. This is somewhat uncharacteristic for me, and, I suppose, for working scientists in general. I've heard the remark (although I forget the source) that the philosophy of science is just about as useful to scientists as ornithology is to birds.

However, at just this time a question has arisen in the United States that will affect the direction of physics research until well into the twenty-first century, and that I think hinges very largely on a philosophical issue. On 30 January of this year the present administration in Washington announced that it had decided to go ahead with the construction of a large new accelerator for elementary particle physics, the Superconducting Supercollider, or SSC for short. 'Large' in this case means that its circumference would be about 53 miles. The circumference is determined by the necessity of accelerating protons to energies of 20 TeV (2×10^{13} electron volt). Within this ring there would travel two counter-rotating beams of protons, that would slam into each other at a number of intersection regions. The intensity of the beams is designed to be such that one would have a collision rate of about one per second for typical processes (with a cross-section of a nanobarn). All of these design parameters lead to a bottom line parameter: the cost in 1986 dollars is estimated to be 4,400 million dollars.

The chief reason for wanting to go ahead with this accelerator is that it would open up a new realm of high energy which we have not yet been able to study. Just as when astronomers start to study the sky at a new wavelength, or when solid state physicists go down another factor of ten in temperature, every time particle accelerators go up a factor of ten in energy we discover exciting new physics. This has generally been the rationale for new accelerators. Occasionally, one can also

point to specific discoveries that can be anticipated from a particular new accelerator. One example is provided by the accelerator built in Berkeley over 30 years ago, the Bevatron, which for the first time was capable of producing particles with masses of 1 GeV. (In those days American

the particle known as the Higgs boson, provided that the Higgs boson is not too heavy. If the Higgs boson is too heavy, then the SSC will discover something else equally interesting.

Let me explain these remarks further. As many people may have heard, there

has been a certain measure of unification among the forces of nature. This unification entails the idea that the symmetry among the forces, specifically the weak nuclear force and the electromagnetic force, is spontaneously broken. It can't be spontaneously broken by the forces we know about, that is, the ordinary strong and weak nuclear forces and the electromagnetic force; therefore there must be a new force in nature which is responsible for the symmetry breaking, like the phonon exchange force in a superconductor. We don't know exactly what that force is. The simplest picture is that it has to do with the existence of a new kind of elementary scalar particle. The members of the multiplet of elementary scalar particles that would be observable as physical particles are called Higgs bosons.

Now, we are not sure that that is actually the correct picture of the mechanism for electroweak symmetry breaking, and we certainly do not know the mass of the Higgs boson. The SSC would be able to discover the



physicists talked about BeV instead of GeV.) The Bevatron was designed to be able to produce antiprotons, and indeed it did so shortly after it went on the air. That was not the only exciting thing done at that accelerator. Quite unexpected was the discovery of a vast forest of new mesonic and baryonic states, that led to a change in our conception of what we mean by an elementary particle. But in planning the Bevatron, it was nice to know in advance that at least one important discovery could be counted on.

The same is true now of the SSC. The SSC is so designed so that it will discover

Higgs boson if its mass is not greater than about 850 GeV, and, of course, if it exists. However, the SSC (to borrow a phrase from M. Chanowitz¹) is a no-lose proposition, because if the Higgs boson does not exist, or is heavier than 850 GeV, there would have to be strong interactions among longitudinally polarized W particles, which the SSC could also discover. These strong interactions would reveal the nature of the spontaneous symmetry breaking between the weak and the electromagnetic interactions.

Now it remains for Congress to decide whether or not to authorize construction

of the accelerator and to appropriate the money. Two committees of the two houses of the Congress, the Committee on Space, Science, and Technology of the House of Representatives and the Subcommittee on Energy Research and Development of the Senate Committee on Energy and Natural Resources, announced hearings on the SSC, both to begin on 7 April of this year. In March, about a month before these hearings, I was asked to testify at them. I may admit that I found this more frightening than inviting. I had been active for some time in working for the building of the SSC, and all this time it had been a nightmare of mine that I would be called up before some tribunal, and asked in a stern voice why it is worth 4.4 billion dollars to find the Higgs boson. Also, I had testified in Congress only once before, and I did not consider myself a master of the art of congressional testimony.

The particle physicists of the United States are in fact quite united behind the idea that this is the right accelerator to build next. (As I said, its purpose is not limited to finding the Higgs boson, which is just one target, but, rather, it is to open up a new range of energies.) But there has been substantial opposition to the SSC from other physicists in the United States. I have read that this is perhaps the most divisive issue that has ever faced American physicists². I believe that in Britain there is a similar debate — not about building an SSC but about whether Britain should remain in CERN, an issue on which I gather not all British scientists agree.

Heavyweights

I knew at the hearings in Washington there would be two heavyweights who would be testifying vigorously against going ahead with the SSC. One would be Philip Anderson, known to everyone as among the leading condensed matter physicists in the world. Anderson has over many years opposed the large sums that are spent on high energy physics. Another to testify would be James Krumhansl, also a distinguished solid state physicist. He, as it happens, taught me physics when I was a freshman at Cornell, but in addition, and this I suspect counts for more, he is slated the year after next to be the president of the American Physical Society.

Both Anderson and Krumhansl I knew would oppose the SSC, and they would be making arguments with which I really couldn't disagree. In particular, I expected that they would argue that money spent on elementary particle physics, high energy physics, whatever you want to call it, is not as sure to yield immediate technological advances as the same money spent on condensed matter physics, and some other fields. I would have to agree with that (though I would

put more emphasis on the benefits of unpredictable discoveries and spin-offs). I expected that they would also argue that elementary particle physics is not more intellectually profound than other areas of physics like, say, condensed matter physics. I would also agree with that. In fact, we've seen in the last few decades a continual trading back and forth of ideas between elementary particle physics and condensed matter physics. We learned about broken symmetry from them, they learned about the renormalization group from us. And now we're all talking about conformal quantum field theories in two dimensions (I don't now who learned that from whom). But it is clear that there's no lack of mathematical profundity in condensed matter physics as compared with elementary particle physics.

The case for spending large sums of money on elementary particle physics has to be made in a different way. It has to be at least in part based on the idea that particle physics (and here, parenthetically, I should say that under 'particle physics' I include quantum field theory, general relativity, and related areas of astrophysics and cosmology) is in *some* sense more fundamental than other areas of physics. This was denied more or less explicitly by Anderson and Krumhansl in their testimony and also by most of the opponents of the SSC. I didn't see how I could avoid this issue in making a case for the SSC. But it's a dangerous argument. It tends to irritate one's friends in other areas of science. Let me give an example, and here I will quote from myself because then I want to quote some comments on my own remarks.

In 1974, shortly after the standard model was put into its final form with the success of quantum chromodynamics, I wrote an article³ for *Scientific American* called "Unified Theories of Elementary Particle Interactions". Just to get the article started I began it with some platitudes, as follows: "One of man's enduring hopes has been to find a few simple general laws that would explain why nature with all its seeming complexity and variety is the way it is. At the present moment the closest we can come to a unified view of nature is a description in terms of elementary particles and their mutual interactions". I really didn't intend to make any important point by this; it was just the sort of thing one says (as, for instance, Einstein: "The supreme test of the physicist is to arrive at those universal elementary laws from which the cosmos can be built up by pure deduction"). Then a decade later I was asked by the MIT Press to review a proposed book, a collection of articles by various scientists. In the manuscript I found an article⁴ by a friend of mine at Harvard, Ernst Mayr, who is one of the most eminent evolutionary biologists of our times. I found that Mayr

cited the remarks in the *Scientific American* article as "a horrible example of the way physicists think". He called me "an uncompromising reductionist".

Agreement

Now, I strongly suspect that there is no real disagreement between Ernst Mayr and myself, and that in fact we are simply talking past each other, and we should try to understand how we agree rather than fight over this. I don't consider myself an uncompromising reductionist. I consider myself a compromising reductionist. I would like to try to formulate in what way elementary particle physics is more fundamental than other areas of physics, trying to narrow this down in such a way that we can all agree on it.

Let me first take up some of the things I don't mean. And here it is useful to look back at some more of Ernst Mayr's writing, because he is in fact the leading opponent of the reductionist tendency within biology, as well as in science in general. He wrote a book⁵ in 1982, *The Growth of Biological Thought*, that contains a well-known attack on reductionism, and so I looked at it to see what Mayr thought reductionism was, and whether or not I consider myself, in his terms, a reductionist.

The first kind of reductionism that Mayr opposes is called by him "theory reductionism". As far as I can understand it, it's the notion that the other sciences will eventually lose their autonomy and all be absorbed into elementary particle physics; they will all be seen as just branches of elementary particle physics.

Now I certainly don't believe that. Even within physics itself, leaving aside biology, we certainly don't look forward to the extinction of thermodynamics and hydrodynamics as separate sciences; we don't even imagine that they are going to be reduced to molecular physics, much less to elementary particle physics. After all, even if you knew everything about water molecules and you had a computer good enough to follow how every molecule in a glass of water moved in space, all you would have would be a mountain of computer tape. How in that mountain of computer tape would you ever recognize the properties that interest you about the water, properties like vorticity, turbulence, entropy and temperature?

There is in the philosophical literature a term, emergence, that is used to describe how, as one goes to higher and higher levels of organization, new concepts emerge that are needed to understand the behaviour at that level. Anderson summarized this neatly in the title of an interesting article⁶ in *Science* in 1972: "More is Different".

Another kind of reductionism is called by Mayr "explanatory reductionism". As I understand it, it is the idea that progress at

the smallest level, say the level of elementary particle physics, is needed to make progress in other sciences, like hydrodynamics, condensed matter physics and so on.

I don't believe that either. I think we probably know all we need to know about elementary particle physics for the purposes of the solid state physicist, for instance, and the biologist. Mayr in his book makes a point that surprised me (but I suppose it's true; he knows a lot more about this than I do), that even the discovery of DNA was not really of much value in the science of transmission genetics. Mayr writes, "To be sure the chemical nature of a number of black boxes in the classical genetic theory were filled in by the discovery of DNA, RNA, and others, but this did not affect in any way the nature of transmission genetics".

I don't disagree with any of this, but it seems to me that in their attacks on reductionism, Mayr and also physicists like Anderson, Krumhansl and others, are missing the point. In fact, we all do have a sense that there are different levels of fundamentality. For instance, even Anderson⁷ calls DNA the "secret of life". We do have a feeling that DNA is fundamental to biology. It's not that it's needed to explain transmission genetics, and it's certainly not needed to explain human behaviour, but DNA is fundamental nonetheless. What is it then about the discovery of DNA that was fundamental to biology? And what is it about particle physics that is fundamental to everything?

Having spoken at length about what I don't mean, now I want to say what I do mean. But I'm not trying here to say anything new, that you don't all already know. What I'm trying to do is precisely the opposite: to identify what we can all agree on.

In all branches of science we try to discover generalizations about nature, and having discovered them we always ask why are they true. I don't mean why we believe they are true, but why they *are* true. Why is nature that way? When we answer this question the answer is always found partly in contingencies, that is, partly in just the nature of the problem that we pose, but partly in other generalizations. And so there is a sense of direction in science, that some generalizations are 'explained' by others.

To take an example relative to the tercentenary celebration of the *Principia*: Kepler made generalizations about planetary motion, Newton made generalizations about the force of gravity and the laws of mechanics. There is no doubt that historically Kepler came first and that Newton, and also Halley and Wren and others, derived the inverse square law of gravity from Kepler's laws. In formal logic, since Kepler's laws and Newton's laws are both true, either one can be said

to imply the other. (After all, in formal logic the statement 'A implies B' just means that it never happens that A is true and B isn't, but if A and B are both true then you can say that A implies B and B implies A.)

Intuition

Nevertheless, quite apart from formal logic, and quite apart from history, we intuitively understand that Newton's laws of motion and law of gravity are more fundamental than Kepler's laws of planetary motion. I don't know exactly what I mean by that; presumably it has something to do with the greater generality of Newton's laws, but about this also it's hard to be precise. But we all know what we mean when we say that Newton's laws 'explain' Kepler's. We probably could use help from professional philosophers in formulating exactly what that statement means, but I do want to be clear that it is a statement about the way the Universe is, not about the way physicists behave. In the same way, even though new concepts 'emerge' when we deal with fluids or many-body systems, we understand perfectly well that hydrodynamics and thermodynamics are what they are because of the principles of microscopic physics. No one thinks that the phenomena of phase transitions and chaos (to take two examples quoted by Krumhansl) could have been understood on the basis of atomic physics without creative new scientific ideas, but does anyone doubt that real materials exhibit these phenomena because of the properties of the particles of which the materials are composed?

Another complication in trying to pin down the elusive concept of 'explanation' is that very often the 'explanations' are only in principle. If you know Newton's laws of motion and the inverse square law of gravity you can deduce Kepler's laws — that's not so hard. On the other hand, we also would say that chemical behaviour, the way molecules behave chemically, is explained by quantum mechanics and Coulomb's law, but we don't really deduce chemical behaviour for very complex

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Hunting the Higgs — computer simulation of a proton-proton collision in the SSC. The picture is taken from "To the Heart of Matter", issued by the Universities Research Association.

molecules that way. We can for simple molecules; we can explain the way two hydrogen atoms interact to form a hydrogen molecule by solving Schrödinger's equation, and these methods can be extended to fairly large molecules, but we can't work out the chemical behaviour of DNA by solving Schrödinger's equation. In this case we can at least fall back on the remark that although we don't in fact calculate the chemical behaviour of such complicated molecules from quantum mechanics and Coulomb's law, we could if we wanted to. We have an algorithm, the variational principle, which is capable of allowing us to calculate anything in chemistry as long as we had a big enough computer and were willing to wait long enough.

The meaning of 'explanation' is even less clear in the case of nuclear behaviour. No one knows how to calculate the spectrum of the iron nucleus, or the way the uranium nucleus behaves when fissioning, from quantum chromodynamics. We don't even have an algorithm; even with the biggest computer imaginable and all the computer time you wanted, we would not today know how to do such calculations. Nevertheless, most of us are convinced that quantum chromodynamics does explain the way nuclei behave. We say it explains it 'in principle', but I am not really sure of what we mean by that.

Still, relying on this intuitive idea that different scientific generalizations explain others, we have a sense of direction in science. There are arrows of scientific explanation, that thread through the space of all scientific generalizations. Having discovered many of these arrows, we can now look at the pattern that has emerged, and we notice a remarkable thing: perhaps the greatest scientific discovery of all. These arrows seem to converge to a common source! Start anywhere in science and, like an unpleasant child, keep asking "Why?". You will eventually get down to the level of the very small.

By the mid-1920s, the arrows of explanation had been traced down to the level of the quantum mechanics of electrons,

photons, atomic nuclei and, standing somewhat off in the corner, the classical theory of gravity. By the 1970s we had reached a deeper level — a quantum field theory of quarks, leptons and gauge bosons known as the standard model, and with gravity still somewhat isolated, described by a not very satisfactory quantum field theory of gravitons. The next step, many of us think, is the theory of superstrings, still under development. I myself, although a late-comer to this field, confess my enthusiasm for it. I think it provides our best hope of making the next step beyond the standard model.

Objective reductionism

Now reductionism, as I've described it in terms of the convergence of arrows of explanation, is not a fact about scientific programmes, but is a fact about nature. I suppose if I had to give a name for it, I could call it objective reductionism. It is very far from a truism. In particular, these arrows of explanation might have led to many different sources. I think it's important to emphasize that, until very recently, most scientists thought that that was the case; this discovery, that the arrows of explanation point down to a common source, is quite new. (In a comment on an earlier version of this talk, Ernst Mayr informs me that what I call 'objective reductionism' is what he means by 'theory reductionism'. Maybe so, but I prefer to keep the separate terms, because I wish to emphasize that what I am talking about here is not the future organization of the human scientific enterprise, but an order inherent in nature itself.)

To underscore this point, I'd like to mention a few examples of the contrary view surviving until well into the twentieth century. The first is biological vitalism, the idea that the usual rules of physics and chemistry need to be modified when applied to living organisms. One might have thought that this idea would have been killed off by the rise of organic chemistry and evolutionary biology in the nineteenth century. However, Max Perutz in his talk at the Schrödinger centenary in London in April reminded us that both Niels Bohr and Erwin Schrödinger believed that the laws of physics as then understood in the 1920s and 1930s were inadequate for understanding life⁸. Perutz explains that the problem of the orderliness of life that bothered Schrödinger was cleared up by advances in the understanding of enzymatic catalysis. Ernst Mayr was careful in his book to disavow any lingering attachment to vitalism, as follows: "Every biologist is fully aware of the fact that molecular biology has demonstrated decisively that all processes in living organisms can be explained in terms of physics and chemistry". (Mayr, by the way, is using the word 'explained' in exactly the same sense as I am here.)

A second example. Lord Kelvin, in a speech to the British Association for the Advancement of Science, around 1900, said⁹, "There is nothing new to be discovered in physics now. All that remains is more and more precise measurement". There is a similar remark of Michelson's that is often quoted¹⁰. These remarks of Kelvin's and Michelson's are usually cited as examples of scientific arrogance and blindness, but I think this is based on a wrong interpretation of what Kelvin and Michelson meant. The reason that Kelvin and Michelson made these remarkable statements is, I would guess, that they had a very narrow idea of what physics was. According to their idea, the subject matter of physics is motion, electricity, magnetism, light and heat, but not much else. They felt that that kind of physics was coming to an end, and in a sense it really was. Kelvin could not possibly have thought in 1900 that physics had already explained chemical behaviour. He didn't think so, but he also didn't think that was a task for physics. He thought that physics and chemistry were sciences on the same level of fundamentalness. We don't think that way today, but it isn't long ago that physicists did think that way.

I said that these arrows of explanation could have led down to a number of separate sciences. They also could have gone around in a circle. This is still a possibility. There is an idea that's not quite dead among physicists and cosmologists, the 'anthropic principle', according to which there are constants of nature whose value is inexplicable except through the observation that if the constants had values other than what they have the Universe would be so different that scientists would not be there to ask their questions. If the anthropic principle were true, there would be a kind of circularity built into nature, and one would then I suppose have to say that there is no one fundamental level — that the arrows of explanation go round in circles. I think most physicists would regard the anthropic principle as a disappointing last resort to fall back on only if we persistently fail to explain the constants of nature and the other properties of nature in a purely microscopic way. We'll just have to see.

Now although what I have called objective reductionism became part of the general scientific understanding only relatively recently (after the development of quantum mechanics in the 1920s), its roots can be traced back to Newton (who else?). Newton was the first to show the possibility of an understanding nature that was both comprehensive and quantitative. Others before him, from Thales to Descartes, had tried to make comprehensive statements about nature, but none of them took up the challenge of explaining actual observations quantitatively in a comprehensive physical theory.

I don't know of any place where Newton lays out this reductionist programme explicitly. The closest I can come to it is a remark in the Preface to the first edition of the *Principia*, written in May 1686. Newton says, "I wish we could derive the rest of the phenomena of nature by the same kind of reasoning from mechanical principles [I suppose he means as in the *Principia*] for I am induced by many reasons to suspect that they may all depend on certain forces". I suppose that the most dramatic example of the opening up by Newton of the possibility of a comprehensive quantitative understanding of nature is in the third book of the *Principia* where Newton reasons that the moon is 60 times further away from the centre of the Earth than Cambridge is (either Cambridge) and therefore the acceleration of the Moon towards the Earth should be less than the acceleration of an apple in Cambridge by a factor of 60². With this argument Newton unites celestial mechanics and observations of falling fruits in a way that I think captures for the first time the enormous power of mathematical reasoning to explain not only idealized systems like planets moving in their orbits, but ultimately everything.

A digression. Since I have been talking about Newton, and also talking about the SSC, a prime example of 'big science', I can't resist remarking that Newton himself was involved in big science¹¹. In 1710, as President of the Royal Society, Newton by royal command was given control of observations at the largest national laboratory for science then in existence in England, the Greenwich Observatory. He was also given the responsibility of overseeing the repair of scientific instruments by the Master of Ordnance, an interesting connection with the military. (This arrangement, incidentally, infuriated the then Astronomer Royal, Flamsteed.)

Gaps

There are many gaps, of course, and perhaps there always will be many gaps in what I have called the chains of explanation. The great moments in the history of science are when these gaps are filled in, as for example when Darwin and Wallace explained how living things, with all their adaptations to their environment, could develop without any continuing external intervention. But there are still gaps.

Also, sometimes it isn't so clear which way the arrows of explanation point. Here's one example, a small one, but one that has bothered me for many years. We know mathematically that as a consequence of Einstein's general theory of relativity gravitational waves should be waves of spin two, and therefore when quantized, the theory of gravity should have in it particles of mass zero and spin two. On the other hand, we also know that any particles of mass zero and spin two

must behave as described by Einstein's general theory of relativity. The question is, which is the explanation of which? Which is more fundamental, general relativity or the existence of particles of mass zero and spin two? I've oscillated in my thinking about this for many years. At the present moment in string theory the fact that the graviton has mass zero and spin two appears as an immediate consequence of the symmetries of the string theory, and the fact that gravity is described by the formalism of Riemannian geometry and general relativity is a somewhat secondary fact, which arises in a way that is still rather mysterious. But I don't know if that is the final answer. I mention this example just to show that although we don't always know which truths are more fundamental, it's still a worthwhile question to ask, because it is a question about the logical order of nature.

I believe that objective reductionism, reductionism as a statement about the convergence of arrows of explanation in nature, is by now ingrained among scientists, not only among physicists but also among biologists like Ernst Mayr. Let me give an example. Here's a quote from the presidential address of Richard Owen to the British Association in 1858 (ref. 12). Owen was an anatomist, generally regarded as the foremost of his time, and a great adversary of Darwin. In his address, Owen says, "Perhaps the most important and significant result of palaeontological research has been the establishment of the axiom of the continuous operation of the ordained becoming of living things". I'm not too clear what precisely Owen means by this axiom. But my point is that today no biologist would make such a statement, even if he or she knew what the axiom meant, because no biologist today would be content with an axiom about biological behaviour that could not be imagined to have an explanation at a more fundamental level. That more fundamental level would have to be the level of physics and chemistry, and the contingency that the Earth is billions of years old. In this sense, we are all reductionists today.

Now, these reflections don't in themselves settle the question of whether the SSC is worth 4.4 billion dollars. In fact, this might be a difficult problem, if we were simply presented with a choice between 4.4 billion dollars spent on the SSC and 4.4 billion dollars spent on other areas of scientific research. However I don't think that that's likely to be the choice with which we are presented. There is evidence that spending on 'big science' tends to increase spending on other science, rather than the reverse. We don't really know with what the SSC will compete for funds. In any case, I haven't tried here to settle the question of whether or not the SSC should be built for 4.4 billion dollars — it is a complicated question, with many side

arguments. All I have intended to argue here is that when the various scientists present their credentials for public support, credentials like practical values, spinoff etc., there is one special credential of elementary particle physics that should be taken into account and treated with respect, and that is that it deals with nature on a level closer to the source of the arrows of explanation than other areas of physics. But how much do you weigh this? That's a matter of taste and judgement, and I'm not paid to make that final decision. However I would like to throw into the balance one more point in favour of the SSC.

"There is one clue in today's elementary particle physics that we are not only at the deepest level we can get right now, but we are at a level which is in fact in absolute terms quite deep, perhaps close to the final source."

I have remarked that the arrows of explanation seem to converge to a common source, and in our work on elementary particle physics we think we're approaching that source. There is one clue in today's elementary particle physics that we are not only at the deepest level we can get right now, but we are at a level which is in fact in absolute terms quite deep, perhaps close to the final source. And here again I would like to quote from myself, from my own testimony in Congress, because afterwards I am going to quote some comments on these remarks, and I want you to know what it is that the comments were about:

There is reason to believe that in elementary particle physics we are learning something about the logical structure of the Universe at a very very deep level. The reason I say this is because as we have been going to higher and higher energies and as we have been studying structures that are smaller and smaller we have found that the laws, the physical principles, that describe what we learn become simpler and simpler. I am not saying that the mathematics gets easier, Lord knows it doesn't. I am not saying that we always find fewer particles in our list of elementary particles. What I am saying is that the rules that we have discovered become increasingly coherent and universal. We are beginning to suspect that this isn't an accident, that it isn't just an accident of the particular problems that we have chosen to study at this moment in the history of physics but there is simplicity, a beauty, that we are finding in the rules that govern matter that mirrors something that is built into the logical structure of the Universe at a very deep level. I think that this kind of discovery is something that is going on in our present civilization at which future men and women and not just physicists will look back with respect.

After I made these remarks there were remarks by other witnesses, and then there were questions from members of the Committee on Space, Science, and Technology. I am going to quote from the

remarks of two of them. The first is Harris W. Fawell, Republican congressman from Illinois. Fawell throughout his questioning had been generally favourable to the SSC. The second is representative Don Ritter, of Pennsylvania, also a Republican, who had been the congressman most opposed to the SSC throughout the morning. (I suppose you could regard this as a modern dialogue between Sagredo and Simplicio.) I quote here from the unedited transcript of the hearings.

Mr Fawell: Thank you very much. I appreciate the testimony of all of you. I think it was excellent. If ever I would want to explain to one and all the reasons why the SSC is needed I am sure I can go to your testimony. It would be very helpful. I wish sometimes we have some one word that could say it all and that is kind of impossible. I guess perhaps Dr Weinberg you came a little close to it and I'm not sure but I took this down. You said you suspect that it isn't all an accident that there are rules which govern matter and I jotted down, will this make us find God? I'm sure you didn't make that claim, but it certainly will enable us to understand so much more about the universe?

Mr Ritter: Will the gentleman yield on that? [That's something congressmen say to each other.] If the gentleman would yield for a moment I would say. . .

Mr Fawell: I'm not sure I want to.

Mr Ritter: If this machine does that I am going to come round and support it.

Now while this dialogue was going on I thought of a number of marvellous observations that I could make to score points for the SSC. However, by the time Mr Ritter reached his final remark I had decided to keep my mouth shut. And that, my friends, is what I learned about the art of congressional testimony. □

1. Chanowitz, M.S. presented at the 23rd International Conference on High Energy Physics, Berkeley, California, July 16-23, 1986 (Lawrence Berkeley Laboratory Publication No. 21973).
2. Dixon, B. *The Scientist* June 15, p.13 (1987).
3. Weinberg, S. *Scientific American* 231, 50 (1974).
4. Mayr, E. in *Evolution at a Crossroads* (ed. Depew, D.J. & Weber, B.H.) (MIT Press, Cambridge, 1985).
5. Mayr, E. *The Growth of Biological Thought* 58-66 (Harvard University Press, Cambridge, 1982).
6. Anderson, P. *Science* 177, 393 (1972).
7. Anderson, P. Letter to *New York Times* June 8 (1987).
8. Perutz, M. in *Schrödinger, Centenary Celebration of a Polymath* (ed. Kilmister, C.W.) 234 (Cambridge University Press, Cambridge, 1987).
9. Salam, A. in an address to the symposium *The Challenge of Higher Energies*, Oxford (1982).
10. Michelson, A.A. *Light Waves and Their Uses* (1903).
11. Mark, H. *Navigation* 26, 25 (1979).
12. *Edinburgh Review* 11, 487-532 (1960); see also Hull, D.L. in *Darwin and his Critics* (Harvard University Press, Cambridge, 1973).

I was greatly aided in preparing the talk at Cambridge and thus this article by conversations with G. Holton, H. Mark, E. Mayr and E. Mendelsohn. I am also grateful for helpful comments on an earlier written version by J. Krumhansl, E.L. Goldwasser, E. Mayr, M. Perutz and S. Wojicki.

Steven Weinberg is in the Theory Group, Department of Physics, University of Texas, Austin, Texas 78712, USA.

GUIDE TO AUTHORS

Please follow these guidelines so that your manuscript may be handled expeditiously.

Nature is an international journal covering all the sciences. Contributors should therefore bear in mind those readers for whom English is a second language and those who work in other fields. Please write clearly and simply, avoiding unnecessary technical terminology. *Nature's* staff will edit manuscripts to those ends if necessary. Contributors should check their proofs carefully.

Because of the competition for space, many of the papers submitted for publication cannot be accepted. For this reason, and because brevity is a great assistance to readers, papers should be as brief as is consistent with intelligibility. Please note that one printed page of *Nature*, without diagrams or other interruptions of the text, has fewer than 1,300 words.

CATEGORIES OF PAPER

Review Articles survey recent developments in a field. Most are commissioned, but suggestions are welcome in the form of a one-page synopsis addressed to the Reviews Coordinator. Length is negotiable in advance but must not exceed six pages of *Nature*.

Articles are research reports whose conclusions are of general interest and which are sufficiently rounded to be a substantial advance in understanding. They should not have more than 3,000 words of text or more than six display items (figures and tables) and should not occupy more than five pages of *Nature*.

Articles should be accompanied by a heading of 50–80 words written to advertise their contents in general terms, to which editors will pay particular attention. A heading (printed in italic type) is not an abstract and should not usually contain numbers or measurements. The study should be introduced in more detail in the first two or three paragraphs, which should also briefly summarize its results and implications.

Articles may contain a few subheadings of two or three words. The meaning of the text should not depend on the subheadings, whose function is to break up the text and to point to what follows. There should be fewer than 50 references.

Letters to Nature are short reports of outstanding novel findings whose implications are general and important enough to be of interest to those outside the field. Letters should not have more than 1,000 words of text or more than four display items and should not occupy more than two pages of *Nature*. The first paragraph should describe, in not more than 150 words, the origins and chief conclusions of the study. Letters should not have subheadings or more than 30 references.

Commentary articles deal with issues in, or arising from, research that are also of interest to readers outside research. Some are commissioned, most are unsolicited. They are normally between one and four pages of *Nature* in length.

News and Views articles are intended to inform non-specialist readers about a recently published advance. Suggestions should be made to the News and Views Editor. Illustrations are welcome. Proposals for meeting reports should be agreed in advance.

Scientific Correspondence is for the discussion of scientific matters, including contributions published in *Nature*. Priority is given to contributions of less than 500 words and five references. Figures are welcome.

GENERAL

All contributions submitted for publication in *Nature* should conform with these rules.

Manuscripts should be typed, double-spaced, with a good-quality printer, on one side of the paper only. Four copies are required, each accompanied by lettered artwork. Four copies of half-tones should be included. Reference lists, figure legends and so on should be on separate sheets, all of which should be double-spaced and numbered. Copies of relevant manuscripts in press or submitted for publication elsewhere should be included, clearly marked as such. Revised and resubmitted manuscripts should also be clearly marked as such and labelled with their reference numbers.

Titles should say what the paper is about with the minimum of technical terminology and in fewer than 80 characters. Authors should avoid active verbs, numerical values, abbreviations and punctuation and should include one or two key words for indexing purposes.

Artwork should be marked individually and clearly with the author's name, figure number and, when known, manuscript reference number. Original artwork should be unlettered. Ideally, no figure should be larger than 28 by 22 cm. Figures with several parts are permitted only if the parts are closely related, either experimentally or logically. Suggestions for cover illustrations, with captions, are welcome. Original artwork (and one copy of the manuscript) will be returned when a manuscript cannot be published.

Figure legends must not exceed 300 words and ideally should be much shorter, and should use telegraphic form wherever possible. The figure should be described first, then, briefly, the method. Reference to a method described elsewhere is preferable to a full description. Methods should not be described in detail in the text.

References should be numbered sequentially as they appear in the text, followed by those in tables and finally those in the figure legends. Reference numbers apply only to papers published or in the press, and a different number should be given to each paper cited. All other forms of reference (including unrefereed abstracts) should be included in the text as personal communication, manuscript in preparation or preprint (with number and institution where appropriate). Text should not be included in the reference list. References should be abbreviated according to the *World List of Scientific Periodicals* (fourth edition, Butterworth, London, 1963–65). The first and last page numbers should be cited. Reference to books should clearly indicate the publisher and the date and place of publication.

Abbreviations, symbols, units and Greek letters should be identified the first time they are used. Acronyms should be avoided as much as possible and, when used, defined. Editors will shorten words if necessary. In case of doubt, SI units should be used.

Footnotes should not be used except for changed addresses or to identify the corresponding author if different from the first-named.

Acknowledgements should be brief. Grant numbers and contribution numbers are not allowed.

Submission. All manuscripts may be submitted to either London or Washington. They should not be addressed to editors by name. Manuscripts or proofs sent by air courier to London should be declared as 'manuscripts' and 'value \$5' to prevent the imposition of import duty and value-added tax (VAT).

Magmatism at rifted continental margins

Robert S. White*, George D. Spence*, Susan R. Fowler*, Dan P. McKenzie*,
Graham K. Westbrook† & Adrian N. Bowen‡

* Bullard Laboratories, Department of Earth Sciences, Madingley Road, Cambridge CB3 0EZ, UK

† Department of Geological Sciences, Birmingham University, PO Box 363, Birmingham B15 2TT, UK

‡ Department of Geological Sciences, Durham University, South Road, Durham DH1 3LE, UK

The intense volcanism and uplift observed on many rifted continental margins, forming basaltic seaward-dipping reflector sequences, is accompanied by the emplacement of a thick igneous section at depth. Partial melting by decompression of passively upwelling asthenosphere that is hotter than normal because it is near a hotspot explains both the thickness of the new igneous crust and the initial elevation of the rifted margins.

WHEN continental lithosphere is stretched it responds by thinning. If the stretching stops before the lithosphere has been reduced to less than about half its original thickness, the thinned region subsides to form a sedimentary basin, as has happened, for example, in the North Sea^{1,2}. Continued stretching causes the continental lithosphere to break, and new oceanic lithosphere forms between the rifted continental plates. If formation of oceanic lithosphere by seafloor spreading continues, an oceanic basin is generated, bordered on either side by rifted continental margins.

Recent seismic and drilling studies of rifted continental margins have shown that some are characterized by extensive volcanic activity during the initial rift stage³⁻⁵, whereas others are apparently devoid of such volcanism⁶. The volcanic margins have remained close to sea level during the rift stage, whereas the non-volcanic margins have subsided during stretching and rifting, typically by ~2 km.

We show that these marked differences in response, in both the amount of magmatism and the occurrence of uplift or subsidence, can be explained by a simple model of partial melting by decompression of asthenosphere as it upwells passively beneath the rift. The only difference between volcanic and non-volcanic margins is that the potential temperature of the asthenosphere is up to 100–150 °C higher beneath the former. (The potential temperature is the temperature that the mantle would have if it moved to the Earth's surface adiabatically with no melting.) We illustrate our model using new results from a volcanic rifted margin, the Hatton Bank margin off north-west Britain (Fig. 1).

The North Atlantic contains both volcanic and non-volcanic types of continental margin. To the north are volcanic margins off Rockall Bank, the Iceland–Faeroes Rise, the Norwegian margin, Jan Mayen and Greenland. To the south the continental margins, exemplified by the Biscay margin of France, are devoid of syn-rift volcanism and the adjacent sea floor is deeper (Fig. 1).

The occurrence of large amounts of volcanic rock on some rifted margins was postulated by Hinz⁷, who interpreted sequences of seaward-dipping reflectors on seismic reflection profiles as caused by stacked layers of subaerial or shallow-marine basalt flows. For some time it was debated whether the seaward-dipping reflectors were indeed volcanic, or were sedimentary deltaic deposits⁸. It is now clear that they are basaltic, both from direct sampling by drilling^{3,9} and from seismic velocity measurements^{10,11}. Recently the debate has centred around whether the dipping reflectors overlie continental^{3,7,12,13} or oceanic^{10,14} crust. We suggest that the question is partly one of terminology. As the young igneous component of the crust must increase oceanwards, different definitions of oceanic and continental crust will lead to disagreement about where the boundary is between the two. It is probable that the rifted margins will contain fragments

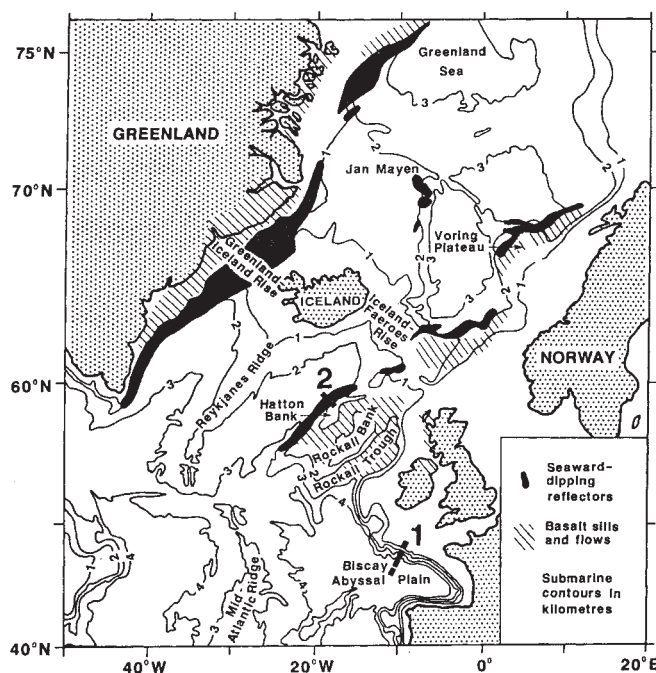


Fig. 1 Distribution of the basalt flows that form prominent seaward-dipping reflector sequences, and contemporaneous basalt sills and flows on the rifted continental margins of the northern North Atlantic. Data compiled from refs 3, 11, 12 and 46–50.

of continental rocks, surrounded by young basalt, as seen in the currently active Afar rift at the southern end of the Red Sea. The magma that is eventually emplaced as basalt flows forming seaward-dipping reflectors is generated by stretching and thinning the continental lithosphere; the exact location at which the melt penetrates to the surface is of little import when considering its cause.

In what follows we will first present the results of recent work¹¹ using both normal-incidence and wide-angle two-ship multi-channel seismic profiles to constrain the structure of the continent–ocean transition across the Hatton Bank margin off north-west Britain. We will then invoke a quantitative model of melt generation in abnormally hot asthenosphere to explain these results.

Hatton Bank margin

The Hatton Bank margin, which has well developed seaward-dipping reflectors, was chosen for study because the continental fragment of Hatton Bank has been isolated from terrigenous sediment supply since it was separated from the European continent in the late Mesozoic¹⁵. As a result, the structure of

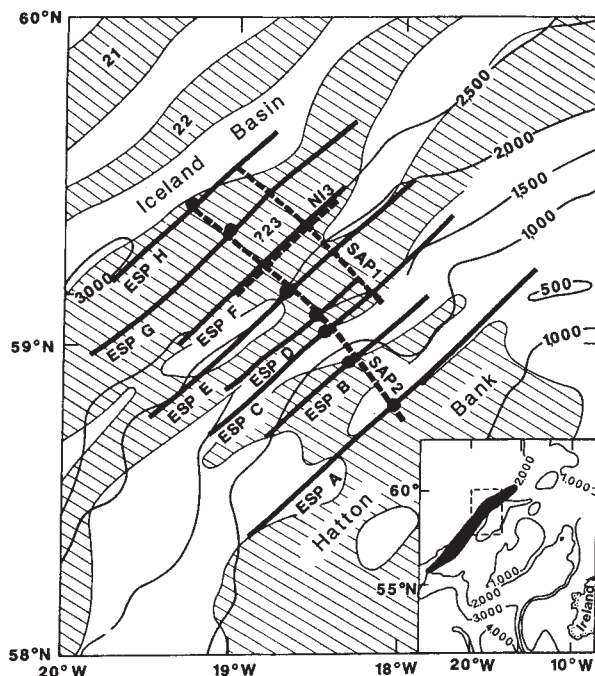


Fig. 2 Location of two-ship multichannel seismic lines discussed in this paper. Shaded areas are positive magnetic anomalies contoured from our data and ref. 51, with reversal number identifications. Dots show midpoints of expanding-spread profiles (ESPs). Bathymetric contours in metres.

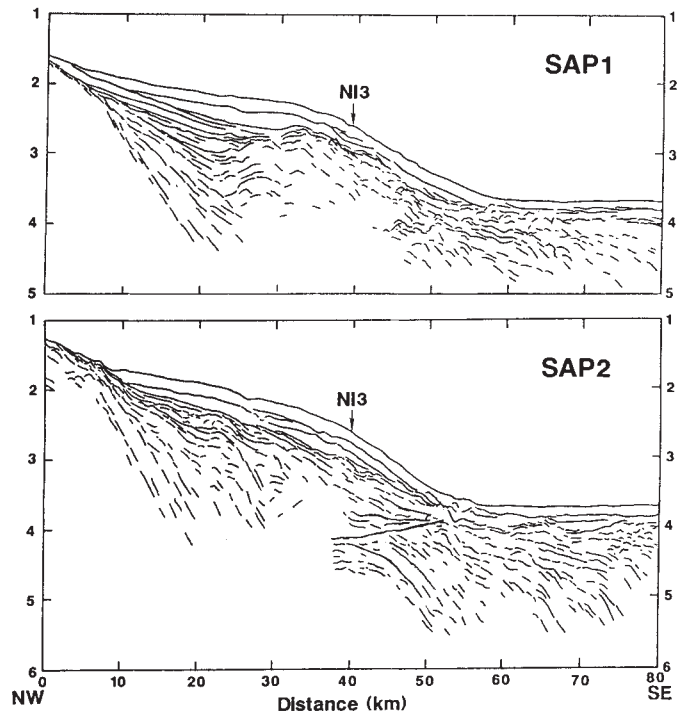
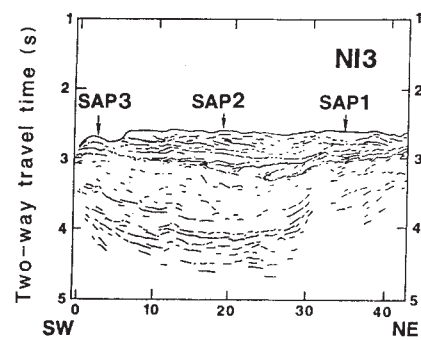


Fig. 3 Line drawings of unmigrated time sections from multichannel seismic reflection profiles SAP1 and SAP2 across the Hatton margin and NI3 parallel to the strike near the foot of the continental slope (see Fig. 2 for location). All phase correlations have been drawn. True vertical exaggeration at the sea floor, $\sim 12:1$; within igneous basement, $\sim 3.5:1$.

the crystalline crust across the continent-ocean transition is not obscured by a thick cover of sediments. There are less than 250 m of unconsolidated post-rift sediments, mainly deposited by northward-flowing contour currents¹⁶.

In 1985 we used the RRS *Discovery* and the RRS *Charles Darwin* to study an 80×100 km area straddling the margin (Fig. 2). Embedded in a grid of 1,400 km of conventional single-ship multi-channel seismic profiles and 7,500 km of magnetics, gravity and bathymetry profiles, were eleven two-ship expanding-spread profiles (ESPs) parallel to the margin and four two-ship synthetic-aperture profiles (SAPs) across the margin: these give excellent control on the configuration and seismic velocities of the igneous crust.

Extrusive igneous rocks

The extrusive igneous section generates two sets of seaward-dipping reflectors in the uppermost 4 km of crust. One set of generally planar reflectors forms a wedge which feathers out on the continental crust of Hatton Bank near the top of the slope (from 5 to 25 km on profiles SAP1 and SAP2, Fig. 3). The other set extends oceanwards beyond the foot of the continental slope into oceanic crust, which exhibits lineated magnetic anomalies caused by seafloor spreading (from 40 to 80 km on SAP1 and SAP2, Fig. 3). The latter set exhibits the characteristic convex-upward configuration described from other volcanic continental margins⁷. Similar dip patterns seen in basalts in Iceland have been explained by the loading effect of flows fed from a retreating vent¹⁷; by analogy, Mutter *et al.*¹⁴ suggested that the seaward-dipping reflectors on the Norwegian margin were formed by "subaerial seafloor spreading".

The lower set of dipping reflectors on the Hatton margin are remarkably uniform along strike (see, for example, the sub-horizontal reflectors below 4 s two-way travel time on NI3, Fig. 3), indicating that the feeder vents were lineated parallel to the margin. Southeastward-dipping reflectors are found on the conjugate Greenland continental margin, and represent lavas extruded on the northwestern side of the feeder vents. The feeder

vents were higher than the surrounding area across which basalts flowed to form the now seaward-dipping reflectors.

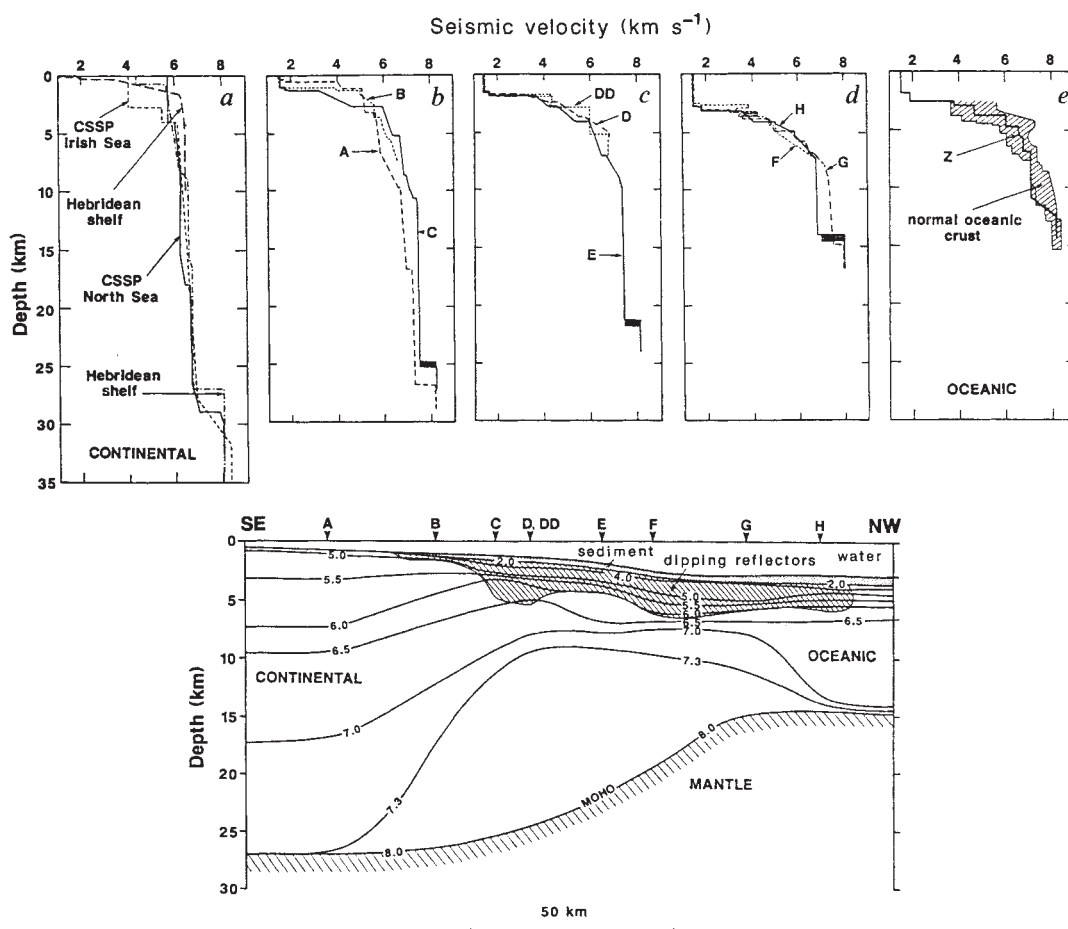
Midway down the slope is a late-stage volcanic pile, ~ 30 km in diameter, which sits unconformably on the dipping reflectors (Fig. 3). The unconformity has been tilted landward by the load. The volcanic pile has a poorly reflective centre and a bumpy upper surface, perhaps representing its original relief.

Crustal structure: igneous accretion

The crust thins from 27 km of continental material beneath Hatton Bank to 12-km-thick oceanic crust beyond the foot of the continental slope¹¹ (Fig. 4). Detailed velocity-depth structures were determined from the ESPs. Corrections for sediment thickness variations were made using coincident multi-channel seismic reflection profiles and ray-tracing computer programs¹⁸. Final velocity-depth profiles shown in Fig. 4 were constrained by modelling the amplitudes of seismic arrivals using reflectivity synthetics¹⁹.

In the extrusive basalt section represented by the dipping reflectors, the velocity increases steeply with depth, from typically 3.5 km s^{-1} at the top to $\sim 6.5 \text{ km s}^{-1}$ at the base. The velocity gradient is probably caused by crack closure with depth, and is

Fig. 4 Lower panel, section across Hatton continental margin along line of profile SAP2 through midpoints of expanding-spread profiles (vertical exaggeration, 2.3:1; Fig. 2 indicates the location). Velocity structure is from the synthetic-seismogram modelling of airgun (profiles B, D, DD, F, H) and explosive (profiles A, C, E, G) ESPs (panels b-d), with iso-velocity contours (in km s^{-1}) drawn through them. Extent of extrusive igneous rocks (shaded) is from the coincident normal-incidence line SAP2 (Fig. 3). Continental velocity-depth profiles constrained by synthetic-seismogram modelling (a) are from Hebridean shelf^{22,23}, Irish Sea and North Sea²⁴; oceanic line Z (e) is from 9-Myr-old crust on the flank of the Reykjanes Ridge³⁴. Shaded area in e, showing the velocity structure of normal oceanic crust, is from North Atlantic profiles on >3-Myr-old crust, compiled in ref. 21.



similar to that found in basalts generated at spreading centres^{20,21}. We attribute the low velocities and local variations in the shallowest structure to the presence of rubble and sediment, perhaps intercalated with lava flows, and to near-surface, possibly subaerial, weathering. Low-velocity layers typically 70–80 m thick within the dipping reflectors probably result from interbedded sediment or debris and from variations and episodicity between basalt flows.

The most significant feature of the lower crust under the continental margin is a prism up to 15 km thick with a well-defined velocity of $7.3\text{--}7.4 \text{ km s}^{-1}$ (Fig. 4). The top of the layer is marked by an abrupt decrease in the velocity gradient and its base by alternating layers of crustal and mantle material over a kilometre-thick transition zone¹¹. The striking feature of the lower-crustal prism is its velocity: nowhere on the adjacent continental crust under the British Isles does the seismic velocity determined by synthetic-seismogram modelling rise above 7.0 km s^{-1} (refs 2, 22–26) (Fig. 4). Support for somewhat higher velocities comes from the LISP results²⁷, based solely on a travel-time analysis which assumed a velocity of “about 7 km s^{-1} ” for the lower crust under northern Britain. But recent synthetic-seismogram modelling of the amplitudes of the LISP data has shown that there is no significant high-velocity material present in the lower crust (P. Barton, personal communication).

As suggested in ref. 11, the high-velocity lower crust is probably produced by igneous material added at the time of rifting, as it is not found in adjacent normal continental sections. From the seismic structure alone we cannot tell whether the lower crust comprises a relatively uniform layer of rock underplated beneath the upper crust or heavily intruded pre-existing material. High lower-crustal velocities of $7.1\text{--}7.6 \text{ km s}^{-1}$ (but most often near 7.2 km s^{-1}) have also been reported from beneath several locations on the eastern margin of North America^{28–33}, and may

in some cases be interpreted as due to igneous underplating³¹.

Across the margin the continental crust thins, and must be heavily intruded by feeder dykes of the overlying extrusive basalts, passing laterally into a region which consists entirely of new igneous rock. We cannot resolve fine details of the lateral transition because the middle crust is hidden beneath the superimposed extrusive basalts. But the velocity structure is consistent with a layer of continental crust that is thinned over a distance of 50–70 km to ~5 km thickness midway down the margin. Near profile ESP C there are 4 km of extrusive basalts above, and up to 15 km of igneous rock beneath, the thinned continental remnant.

Melt production by asthenosphere upwelling

A huge volume of igneous rock was added to the continental margins during rifting. Roberts *et al.*³ have estimated that $>2 \times 10^6 \text{ km}^3$ of basalts were extruded onto the rifting margins of the northern North Atlantic, which is similar to the volume, for example, of the basalts in the Deccan volcanic province. If the evidence of extensive addition of igneous rock to the lower crust is also considered, the total volume of new igneous rock is likely to be as large as $(5\text{--}10) \times 10^6 \text{ km}^3$. Furthermore, the bulk of the basalt was erupted during a very short time, probably only 2 Myr for all of the northern North Atlantic margins.

As the continental lithosphere on a rifting margin is stretched and thinned, so asthenosphere upwells to fill the space. As the asthenosphere rises, it decompresses and generates partial melt. The amount of partial melt produced depends on the potential temperature of the asthenosphere, with more melt being produced when the asthenosphere is hotter, and on the amount of decompression, with more melting occurring where the lithosphere thinning is greatest. Provided the lithosphere thins sufficiently, partial melt can be produced under continental rift

zones even if they do not go on to form new oceanic crust³⁵. Foucher *et al.*³⁶ have shown that the transition from stretched continental to entirely new igneous oceanic crust occurs for stretching factors greater than 3, for a rift over asthenosphere of normal temperature.

Foucher *et al.* adopted a simple analytical expression from ref. 37 which approximates the dependence of partial melting on pressure and temperature in the upper mantle. We have used new relations for melt generation as a function of pressure and temperature, calculated from an accurate fit to all of the available experimental observations⁴⁰. The curves shown in Fig. 5 are calculated from this new relation, although they do not differ greatly from those obtained earlier³⁸ using simpler expressions.

The volume of partial melt generated in the asthenosphere under the thinning lithosphere is calculated by requiring the entropy of the upwelling material to be constant, and ignoring the movement between melt and residue; the calculation therefore assumes that the upwelling is rapid. The total thickness of melt generated is then calculated by integration and is assumed to be added to the crust. The mean potential temperature of the convecting upper mantle must generate the mean oceanic crustal thickness when the upwelling reaches the surface, and must be close to 1,280 °C. A similar normal asthenosphere temperature of 1,261–1,295 °C was calculated by Foucher *et al.*³⁶. The uncertainty in this value is principally due to that in the latent heat of melting, taken here to be $3.9 \times 10^5 \text{ J kg}^{-1}$, and is not likely to exceed 20 °C. If the potential temperature of the upwelling mantle is greater, because it once formed part of an upwelling jet beneath a hotspot, then more melting occurs.

The volume of melt generated depends critically on the potential temperature of the asthenosphere, T_p . When 100-km-thick continental lithosphere is thinned by a factor of 5, which is appropriate for a continental margin, asthenosphere upwelling with $T_p = 1,280^\circ\text{C}$ produces only 2 km of melt, whereas for $T_p = 1,380^\circ\text{C}$, ~8 km of melt is generated and for $T_p = 1,480^\circ\text{C}$, nearly 20 km of melt results (Fig. 5). For the Hatton margin an asthenosphere potential temperature of ~1,400–1,450 °C produces the observed thickness of igneous rock under the rifted margin. This is ~100–150 °C hotter than the mean potential temperature of the asthenosphere beneath oceanic spreading centres. We assume that the melt separates from the residual solid as soon as it is generated and is added to the crust. This behaviour accounts for the outburst of magmatic activity which is contemporaneous with the tectonic stretching, and also is consistent with the bulk of the extrusion occurring very rapidly.

As the lithosphere is thinned by uniform stretching, it subsides to maintain isostatic equilibrium¹, but this subsidence is counterbalanced by uplift resulting from the addition of new igneous material to the crust. Because stretched continental crust has very little flexural rigidity^{2,39}, we can calculate the uplift following stretching and the addition of igneous rock to the crust very simply by assuming that local isostatic equilibrium is maintained. At a normal asthenosphere temperature of 1,280 °C the crust subsides by >2 km at the time of rifting (Fig. 5b), but for higher potential temperatures the addition of igneous rock to the crust considerably reduces the amount of subsidence as it rifts (Fig. 5b). Whether the melt is intruded into the crust or is erupted at the surface probably depends on its density. Its mode of emplacement will not, however, affect the surface elevation.

In addition to the uplift caused by the addition of new igneous rock, considerable dynamic uplift can be superimposed if the high asthenosphere temperatures are caused by advected mantle in a convecting hotspot⁴¹. This additional uplift can reach up to 1 km or more over regions of up to 2,000 km diameter. Thermal expansion of the lithosphere may also contribute to uplift. The net result of these processes—the addition of igneous material to the crust, the dynamic uplift caused by mantle convection and uplift caused by lithosphere heating—is to elevate the crust. In a region above hot asthenosphere, therefore, it is very likely

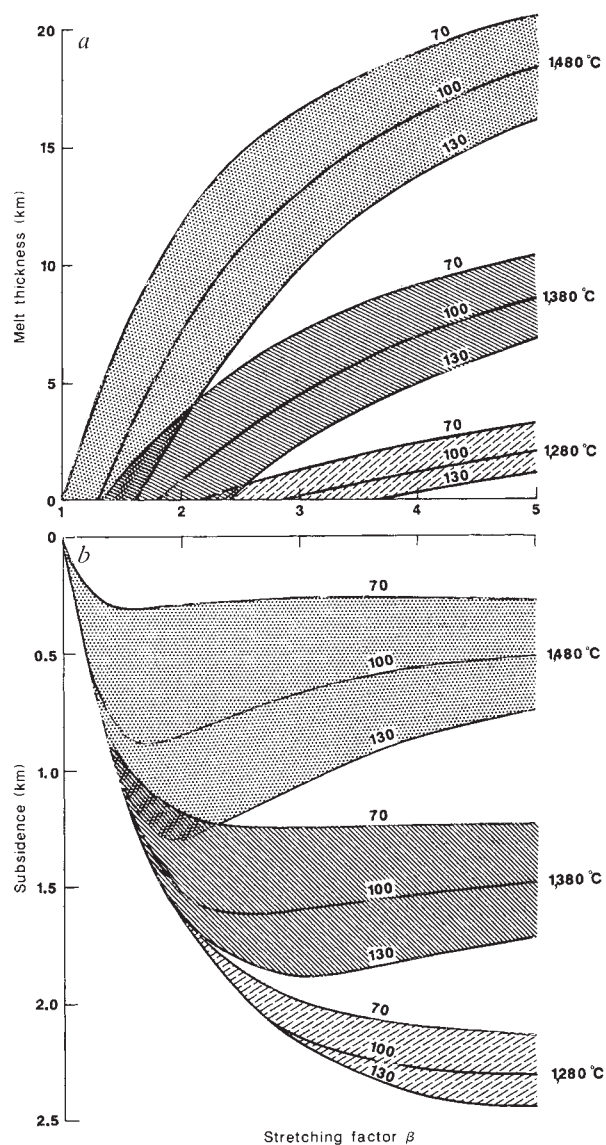


Fig. 5 *a*, Thickness of partial melt generated by decompression of passively upwelling asthenosphere for different amounts of lithosphere thinning, assuming uniform stretching. Curves are shown for three different representative asthenosphere potential temperatures: 1,280, 1,380 and 1,480 °C. For each temperature the curves are shown for three initial lithosphere thicknesses: 70, 100 and 130 km. *b*, Amount of subsidence caused at the time of rifting by the combination of lithosphere thinning due to uniform stretching and crustal thickening caused by the addition of an appropriate volume of new igneous rock calculated from the curves in *a*. The crust is assumed to be in isostatic equilibrium. The density of the igneous rock added to the crust is calculated from the composition of the partial melt⁴⁰, with normative hypersthene distributed between olivine and quartz. The effects of dynamic uplift caused by flow in the asthenosphere and of uplift generated by lithosphere heating are not included.

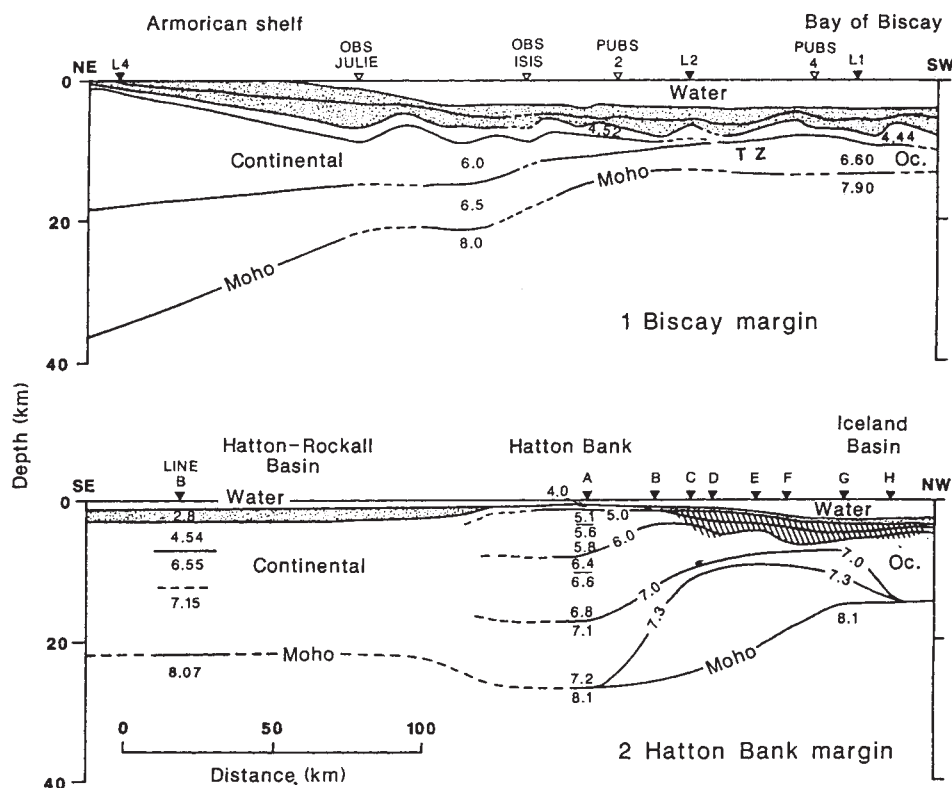
that the rift will remain near or above sea level as it rifts. This is consistent with the observation that the basalts of the dipping reflectors were extruded close to sea level.

Thus, a very simple model of melt production in the passively rising asthenosphere under a continental rift zone can explain the volume of new igneous crust, elevation of the rifted margin above sea level and the rapid outpouring of volcanic rocks contemporaneous with rifting.

Discussion

Our model requires the asthenosphere temperature at the time of rifting to have been elevated by ~100–150 °C across a 2,000-

Fig. 6 Interpreted cross-sections drawn at the same scale across the non-volcanic Biscay continental margin (location 1 on Fig. 1, redrawn from ref. 6) and the volcanic Hatton continental margin (location 2 on Fig. 1). Stipple, syn- and post-rift sediments; TZ, transition zone between continental and oceanic crust from ref. 6; hatching, dipping reflectors. ESPs A-H as in Fig. 4.



km region from the southern end of Rockall Plateau to the Greenland Sea (Fig. 1). A thermal anomaly of this magnitude and extent is in excellent agreement with upper-mantle temperatures beneath currently active hotspots such as the well-modelled Cape Verde Rise⁴¹. The rising jet beneath the Cape Verde volcanic islands is only 100–200 km in diameter, but it is deflected sideways by the overlying plate to form a mushroom of anomalously hot material ~1,500 km across. This generates a bathymetric swell and geoid anomaly with a similar width of ~1,500 km. The swells around other active hotspots have typical dimensions of 1,000–2,000 km, and probably have underlying upper-mantle temperature anomalies similar to those modelled for the Cape Verde Rise⁴¹.

The North Atlantic is dominated by a hotspot which is currently centred beneath Iceland. As the ocean opened, this hotspot left a trace of 20–35-km-thick igneous crust under the Greenland–Iceland–Faeroes Rise⁴² (Fig. 1). Within a 1,000 km radius of Iceland, ocean depths are ~1 km shallower than normally expected for oceanic lithosphere of the same age, owing to the mushroom of hot underlying mantle^{43,44}. When the continent rifted, hot mantle from this region upwelled and partially melted to produce the magmatic outburst.

At the foot of the continental margin the oceanic crust is ~12 km thick. Away from the margin and towards the active oceanic spreading centre of the Reykjanes Ridge, the crustal thickness decreases slightly to ~10 km (ref. 34), which is significantly thicker than the 6–7 km thickness of normal oceanic crust²¹. This is indicative of the continued presence of hot upper mantle associated with the Iceland plume, as is the abnormally shallow bathymetry of the Reykjanes Ridge.

The maximum thickness of the igneous section on the rifted continental margin is >15 km, decreasing to 10–12 km in the adjacent, fully oceanic crust. Several factors may lead to enhanced thicknesses under the rifted margin, compared to the adjacent oceanic crust.

First, the pressure perturbation caused by the separating plates tends to focus the partial melt towards the continental margins

(M. Spiegelman and D. McKenzie, manuscript in preparation). Second, if the lithosphere thinning is not uniform, major detachment faults may localize the lithosphere thinning while allowing broader-scale upper-crustal rifting: this would localize the melt production in the region of greatest lithosphere thinning.

Third, the breaking of the continental lithosphere to form the new ocean basin considerably modified the conditions above the mantle plume now beneath Iceland. Before rifting, the plume was overlain by unbroken continental lithosphere and all of the advected mantle in the central plume was deflected sideways; very little melt penetrated through the continental lithosphere. It was across this region of anomalously hot, advected mantle that the rift occurred, leading to high melt production as it upwelled under the thinning lithosphere. Once the lithosphere was broken, however, partial melt generated in the narrow central plume was able to bleed directly to the surface and produce huge quantities of igneous rock beneath the Iceland–Faeroes Rise. Although much of the residual mantle is still deflected laterally to form the swell seen today around Iceland, the melt extracted beneath Iceland itself removes enormous amounts of heat and depletes the mantle of the least refractory material. So although the remaining mantle tapped beneath the Reykjanes Ridge spreading centre is still hotter than normal, and therefore produced a thicker crust than on spreading ridges elsewhere, it does not produce as much melt as when advected mantle upwelled at the initial stage of continental rifting.

Figure 6 shows a comparison between the structure of the volcanic Hatton margin and that of the Biscay margin^{6,45}, which was away from the influence of any hotspot at the time of rifting. Under stretched margins overlying normal-temperature asthenosphere, such as Biscay, only a few kilometres of melt would be expected to form (ref. 36 and Fig. 5). Under the Hatton margin, with elevated asthenosphere temperatures, huge volumes of melt were generated. An anomalously thick igneous section is also found under the rifted margin of the Vøring Plateau, off Norway⁴⁶ in the North Atlantic (Fig. 1). The Vøring Plateau has a history similar to that of the Hatton margin

described here in detail, and was influenced by elevated asthenosphere temperatures caused by the same Icelandic hot-spot. Our predictions of melt generation also apply to continental rift zones that have not stretched sufficiently to form new oceanic basins, but which nevertheless exhibit extensive underplating and extrusive volcanism: the Basin and Range province is a good example.

Conclusions

A simple quantitative model of partial melt formation and extraction in asthenosphere as it upwells passively under thinned continental lithosphere can explain the major features and subsidence pattern of continental margins. Where the continental rift occurs above a region of normal-temperature asthenosphere

the crust thins and subsides, with only minor addition of igneous material. Where the asthenosphere temperature is increased by 100–150 °C by advected heat, great quantities of partial melt are generated by rapid rifting. The melt separates quickly and may be erupted through or intruded into the crust, depending on its density. In either case the elevation of the surface increases. Basalt flows are tilted by the loading of subsequent lavas as the vents retreat oceanward. The first flows are extruded on to pre-existing continental crust, the later ones on to crust containing steadily increasing amounts of new igneous material.

We thank Research Vessel Services, Barry, for providing two research ships with excellent support, NERC for funding, Barbara Dyson for typing the manuscript and Phyl Fisher for drafting the figures.

Received 23 April; accepted 8 October 1987.

1. McKenzie, D. P. *Earth planet. Sci. Lett.* **40**, 25–32 (1978).
2. Barton, P. & Wood, R. *Geophys. J.R. astr. Soc.* **79**, 987–1022 (1984).
3. Roberts, D. G., Backman, J., Morton, A. C., Murray, J. W. & Keene, J. B. *Init. Rep. DSDP* **81**, 883–911 (1984).
4. Parson, L. M., Masson, D. G., Miles, P. R. & Pelton, C. D. *Inst. oceanogr. Sci. Rep. No.* 223 (1986).
5. Kinsman, D. J. J. in *Petroleum and Global Tectonics* (eds Fischer, A. G. & Judson, S.) 83–126 (Princeton University Press, 1975).
6. Ginzburg, A. et al. *Ann. Geophysicae* **3**, 499–510 (1985).
7. Hinz, K. *Geol. Jb. E22*, 3–28 (1981).
8. Roberts, D. G., Montadert, L. & Searle, R. C. *Init. Rep. DSDP* **48**, 1061–1088 (1979).
9. ODP Leg 104 Scientific Party *Geotimes* March, 12–14 (1986).
10. Mutter, J. C., Talwani, M. & Stoffa, P. L. *J. geophys. Res.* **89**, 483–502.
11. White, R. S. et al. *Geophys. J.R. astr. Soc.* **89**, 265–272 (1987).
12. Smythe, D. K. in *Structure and Development of the Greenland–Scotland Ridge* (eds Bott, M. H. P., Saxov, S., Talwani, M. & Thiede, J.) 109–119 (Plenum, New York, 1983).
13. Morton, A. C. & Taylor, P. N. *Nature* **326**, 381–383 (1987).
14. Mutter, J. C., Talwani, M. & Stoffa, P. L. *Geology* **10**, 353–357 (1982).
15. Srivastava, S. P. & Tapscott, C. R. in *The Geology of North America* Vol. M (eds Vogt, P. R. & Tucholke, B. E.) 379–404 (Geological Society of America, Boulder, 1986).
16. McCave, I. N., Lonsdale, P. F., Hollister, C. D. & Gardner, W. D. *J. sedim. Petrol.* **50**, 1049–1062 (1980).
17. Palmason, G. *J. Geophys.* **47**, 7–18 (1980).
18. Cervený, V., Molotkov, I. A. & Psencik, I. *Ray Method in Seismology* (Univerzita Karlova, Praha, 1977).
19. Kennett, B. L. N. *Geophys. J. R. astr. Soc.* **42**, 579–589 (1975).
20. Whitmarsh, R. B. *Earth planet. Sci. Lett.* **37**, 451–464 (1978).
21. White, R. S. in *Ophiolites and Oceanic Lithosphere* (eds Gass, I. G., Lippard, S. J. & Shelton, A. W.) 34–44 (Geological Society, London, 1984).
22. Hughes, V. J., White, R. S. & Jones, E. J. W. *Ann. Geophysicae* **2**, 669–678 (1984).
23. Powell, C. M. R. & Sinha, M. C. *Geophys. J.R. astr. Soc.* **89**, 259–264 (1987).
24. Bott, M. H. P. et al. *Nature* **314**, 724–727 (1985).
25. Bott, M. H. P., Armour, A. R., Himsworth, E. M., Murphy, T. & Wylie, G. *Tectonophysics* **59**, 217–231 (1979).
26. Smith, P. J. & Bott, M. H. P. *Geophys. J.R. astr. Soc.* **40**, 187–205 (1975).
27. Bamford, D., Nunn, K., Prodehl, C. & Jacob, B. *Geophys. J.R. astr. Soc.* **54**, 43–60 (1978).
28. Austin, J. A. Jr thesis, Woods Hole Oceanographic Inst. (1978).
29. Keen, C. E., Keen, M. J., Barrett, D. L. & Heffler, D. E. in *Offshore Geology of Eastern Canada* (eds Van der Linden, W. J. M. & Wade, J. A.) 189–197 (Geological Survey of Canada, Pap. 74–30, Ottawa, 1975).
30. Sheridan, R. E., Grow, J. A., Behrendt, J. C. & Bayer, K. C. *Tectonophysics* **59**, 1–26 (1979).
31. LASE Study Group *Mar. Petrol. Geol.* **3**, 234–242 (1986).
32. Trehu, A. M. et al. *Eos* **67**, 1103 (1986).
33. Mithal, R. & Diebold, J. A. *Geophys. J.R. astr. Soc.* **89**, 965–986 (1987).
34. Bunch, A. W. H. & Kennett, B. L. N. *Geophys. J.R. astr. Soc.* **61**, 141–166 (1980).
35. Dixon, J. E., Fitton, J. G. & Frost, R. T. C. in *Petroleum Geology of the Continental Shelf of North-West Europe* (eds Illing, L. V. & Hobson, G. D.) 121–137 (Institute of Petroleum, London, 1981).
36. Foucher, J.-P., Le Pichon, X. & Sibuet, J.-C. *Phil. Trans. R. Soc. A* **305**, 27–43 (1982).
37. Ahern, J. L. & Turcotte, D. L. *Earth planet. Sci. Lett.* **45**, 115–122 (1979).
38. McKenzie, D. P. *J. Petrology* **25**, 713–765 (1984).
39. Fowler, S. R. & McKenzie, D. P. *Mar. geophys. Res.* (in the press).
40. McKenzie, D. & Bickle, M. J. *J. Petrology* (in the press).
41. Courtney, R. C. & White, R. S. *Geophys. J.R. astr. Soc.* **87**, 815–867 (1986).
42. Bott, M. H. P. & Gunnarsson, K. J. *Geophys.* **47**, 221–227 (1980).
43. Anderson, R. N., McKenzie, D. P. & Sclater, J. G. *Earth planet. Sci. Lett.* **18**, 391–407 (1973).
44. Haigh, B. I. R. *Geophys. J.R. astr. Soc.* **33**, 405–420 (1973).
45. Whitmarsh, R. B., Avedik, F. & Saunders, M. R. *Geophys. J.R. astr. Soc.* **86**, 589–602 (1986).
46. Mutter, J. C., Buck, W. R. & Zehnder, C. M. *J. geophys. Res.* (in the press).
47. Larsen, H. C. in *Petroleum Geology of the North European Margin* (eds Spencer, A. M. et al.) 329–339 (Graham & Trotman, London, 1984).
48. Nunn, A. G. in *Structure and Development of the Greenland–Scotland Ridge* (eds Bott, M. H. P., Saxov, S., Talwani, M. & Thiede, J.) 11–30 (Plenum, New York, 1983).
49. Skogseid, J. & Eldholm, O. *J. geophys. Res.* (in the press).
50. Uruski, C. I. & Parson, L. M. *Inst. oceanogr. Sci. Rep. No.* 214 (1985).
51. Roberts, D. G. & Jones, M. T. *Magnetic Anomalies in the Northeast Atlantic, North Sheet* (Institute of Oceanographic Sciences, Wormley).

A human retinoic acid receptor which belongs to the family of nuclear receptors

Martin Petkovich, Nigel J. Brand, Andrée Krust & Pierre Chambon

Laboratoire de Génétique Moléculaire des Eucaryotes du CNRS, Unité 184 de Biologie Moléculaire et de Génie Génétique de l'INSERM, Institut de Chimie Biologique, Faculté de Médecine, 11 rue Humann, 67085 Strasbourg-Cédex, France

A cDNA encoding a protein that binds retinoic acid with high affinity has been cloned. The protein is homologous to the receptors for steroid hormones, thyroid hormones and vitamin D₃, and appears to be a retinoic acid-inducible trans-acting enhancer factor, suggesting that the molecular mechanisms of the effect of retinoids (vitamin A) on embryonic development, differentiation and tumour cell growth are similar to those described for other members of this nuclear receptor family.

RETINOL (vitamin A) and its derivative retinoic acid (RA) are essential in the control of epithelial cell growth and cellular differentiation (see refs 1–5) and RA inhibits the growth of some malignant cells (see refs 3 and 6). RA, which cannot be reduced to retinol *in vivo*, may be the active retinoid in most of these processes, although retinol seems to be indispensable in vision and spermatogenesis (see ref. 7). RA also has striking effects on pattern formation in developing and regenerating limbs, and may be the natural morphogen providing the positional information in the postulated morphogenetic gradient of the developing

chick limb bud (see refs 8–12 for reviews).

RA-induced differentiation of F9 mouse teratocarcinoma stem cells into primitive endoderm-like cells is accompanied by increased transcription of specific genes^{2,13–16} and it has been suggested that RA may, like steroid hormones, mediate its transcriptional effects via specific nuclear receptors¹⁷. Nuclear localization of RA has been demonstrated^{1,2,5,18} and specific cellular binding proteins for RA (CRABP^{2,6,7,9,19,20}) and retinol (CRBP^{7,19,21}) have been identified in target tissues. The amino acid sequences of CRABP and CRBP have been established

hER (AA192-199)	5'-GCT	TCA	GGC	TAC	CAT	TAT	GGA	GTC-3'
hGR (AA428-435)	5'-GCT	TCA	GGG	TGT	CAT	TAT	GGA	GTC-3'
hPR (AA574-581)	5'-GCA	TCA	GGC	TGT	CAT	TAT	GGT	GTC-3'
v-ERBA (AA45-52)	5'-GCC	ACC	GGC	TAC	CAC	TAC	CGC	TGC-3'
rGR (AA447-454)	5'-GCT	TCA	GGG	TGT	CAT	TAC	GGG	GTG-3'
cER (AA186-193)	5'-GCT	TCA	GGC	TAC	CAC	TAT	GGG	GTC-3'
Consensus sequence	5'-GC ^A	TCA	GG ^C	TG ^T	CA ^T	TA ^T	GG ^A	GTC-3'
Oligonucleotide probe	3'-CG ^T	AGT	CC ^G	AT ^G	GTG	ATG	CC ^T	CAG-5'
	•	•••	•••	•••	•••	••	•••	•••
hRAR (AA65-72)	5'-TCC	TCA	GGC	TAC	CAC	TAT	GGG	GTC-3'

Fig. 1 A consensus oligonucleotide probe specific for nuclear hormone receptors. DNA sequences corresponding to a highly conserved region of 24 nucleotides within the DNA-binding domain of various nuclear receptors are shown. Numbers on the left are the amino acid positions for the following receptor cDNA sequences: cER, chicken oestrogen receptor²⁶; hER, human oestrogen receptor²⁵; hGR, human glucocorticoid receptor⁴⁰; hPR, human progesterone receptor (ref. 41 and Krust *et al.*, unpublished); v-erba, viral oncogene *erba*⁵⁴ and rGR, rat glucocorticoid receptor⁵⁵. A consensus sequence based on these six sequences is shown underneath. The oligonucleotide probe complementary to this consensus that was used to screen MCF-7 and T47D cDNA libraries is shown below. The corresponding sequence found in the hRAR cDNA is shown at the bottom. Dots, hRAR cDNA nucleotides matching those of the probe.

and their complementary DNA and genes have been cloned (see refs 5 and 22). CRABP has been detected in a variety of cultured cell lines and a number of correlative observations support the proposal that it mediates the action of RA^{2,4,5,9,14}. Binding of RA and retinol to specific nuclear chromatin acceptor sites can be achieved by incubating isolated nuclei with preformed RA-CRABP and retinol-CRBP complexes, respectively^{18,23}. CRABP and CRBP may act as shuttles between the cytoplasm and the nucleus, transferring the retinoids to their putative chromatin acceptor sites, ultimately resulting in modulation of transcription of specific genes²³. In this respect, the proposed mechanism of action of retinoids differs from that of steroid hormones, for which no such shuttle proteins are known (see ref. 17 for review).

The cDNAs of a number of steroid hormone receptors, including the oestrogen (ER), progesterone (PR) and glucocorticoid (GR) receptors have been cloned. These receptors are structurally related and share sequence homology with the avian erythroblastosis virus (AEV) *v-erbA* oncogene²⁴⁻²⁶, the mammalian cellular counterparts of which are nuclear receptors for triiodothyronine (T3), a non-steroidal thyroid hormone^{27,28}. Subsequent studies showed that the vitamin D₃²⁹ and mineralocorticoid³⁰ receptors also belong to this multigene family of nuclear receptors which act as inducible transcriptional enhancer factors (see ref. 17). Two evolutionary-conserved functional domains, termed regions C and E by analogy with the ER (ref. 26; see Fig. 5a), are common to all members of this family. Comparison of the C regions indicates a 50–60% identity over a stretch of 66 or 68 amino acids. It has been shown in the cases of ER and GR that region C, which contains the domain binding to the *cis*-acting hormone-responsive enhancer element, is responsible for the specificity of transcriptional activation of the target genes (see refs 31, 32). Region C may fold into two DNA-binding fingers^{17,33}, each involving two pairs of highly conserved cysteine residues (see ref. 34 for review and Fig. 3a). Region E, which contains the hormone-binding domain^{35,36}, is less well conserved than region C between the various nuclear receptors (see Fig. 5), particularly when comparing steroid and thyroid hormone receptors (ref. 34 and Fig. 3b).

Acting on the assumption that the putative nuclear receptor for RA may also be a member of the nuclear receptor multigene family, we designed a region C consensus oligonucleotide probe and used it to screen cDNA libraries derived from breast cancer cells, known to contain retinoid-binding proteins, whose growth is severely decreased in the presence of physiological concentrations of RA (10⁻⁸ M)^{37,38}.

Cloning of a receptor-like cDNA

A consensus oligonucleotide probe corresponding to a highly conserved sequence in region C of several members of the nuclear receptor family (Fig. 1) was used to screen two 'randomly primed' λ gt cDNA libraries prepared using total poly(A)⁺ RNA from the human breast cancer cell lines MCF-7 and T47D. Phage plaques giving positive signals were re-

screened, first with the same oligonucleotide probe and then with cDNA probes corresponding to region C of human oestrogen (hER)²⁵, progesterone (hPR) (A.K., unpublished data) and glucocorticoid (hGR)³⁹ receptors to eliminate clones of these receptors. The cDNA inserts of the remaining positive clones were subcloned into the plasmid vector pEMBL18⁺ for restriction enzyme mapping and sequencing. Two clones were obtained from the MCF-7 library, one containing a 1.9-kilobase (kb) cDNA insert (clone p63) and the other a 0.6-kb insert that was colinear with the 5' end of the 1.9-kb fragment. The 0.8, 1.1 and 1.3-kb cDNA inserts present in three T47D cDNA clones were also colinear with various parts of the MCF-7 clone p63 (data not shown). All five clones contained the same sequence 5'-TCCTCAGGCTACCACTATGGGGTC-3', which matches 21 out of 24 residues of the probe consensus sequence (Fig. 1). The sequence of the cDNA insert of clone p63 is shown in Fig. 2 from nucleotide 260 to the 3' end. This sequence contains a major open reading frame of 1,296 nucleotides, starting at position 317, that encodes a 432 amino-acid protein (hereafter referred to as hRAR) with a predicted molecular mass (M_r) of 47,682. The existence of this open reading frame was confirmed by inserting the *KpnI-EcoRI* fragment (see Fig. 2) into the Gene-scribe-Z vector pTZ19U, and transcribing *in vitro* with T7 RNA polymerase. The resulting RNA yielded a protein of $M_r \sim 48,000$ (48K) in a reticulocyte lysate translation system (data not shown).

The cDNA-deduced hRAR protein sequence was compared with those of other members of the human nuclear receptor family. This comparison revealed two regions presenting a high degree of similarity (boxed sequences in Fig. 2 and Fig. 3). The first region (region C in Fig. 2 and Fig. 3a) corresponds to the highly conserved 66 to 68 amino-acid sequence present in all nuclear receptors, which may fold into two zinc-stabilized DNA-binding fingers. The 66 amino-acid region C of hRAR is 62%, 59%, 48%, 45% and 43% similar to the corresponding region of the thyroid hormone receptor hc-erbA β (also known as hTR β ^{27,28}), hER²⁵, human mineralocorticoid receptor (hMR)³⁰, hGR⁴⁰ and hPR⁴¹, respectively. The second region (region E in Fig. 2 and Fig. 3b), which is less well-conserved than region C between the various members of the nuclear receptor family, corresponds to the steroid binding domain in the case of ER³⁵, GR^{42,43} and PR³⁶. A comparison of region E of the hRAR with the other receptors indicates the highest degree of similarity with hc-erbA β (34%); a similar degree of identity was found with the *v-erbA* sequence), whereas 15–18% similarity is observed with hER, hGR, hMR or hPR over the same 220 amino-acid sequence. The similarity between the hRAR sequence from amino acid 269 to 386 and the hc-erbA β sequence from amino acid 338 to the C-terminus is particularly noteworthy, because there is a lower degree of similarity with the steroid hormone receptors in this region. This difference in similarity indicates that hRAR and the thyroid hormone receptor hc-erbA β may belong to a distinct sub-group of nuclear receptors which diverged early in evolution from the steroid hormone

Fig. 2 cDNA sequence and deduced amino acid sequence of hRAR. The DNA sequence is numbered on the left according to the sequence of the cDNA clone p63, which contains a 1.9-kb *EcoRI* cDNA insert and is derived from an MCF-7 human breast cancer cDNA library. The protein sequence for the major open reading frame of 1,296 nucleotides (nt) is presented below the DNA sequence, using the standard three letter code for amino acids, and the termination codon TGA is denoted by an asterisk underneath the codon. The protein sequence is numbered on the right-hand side of the figure. A unique *KpnI* site at position 279 is underlined and the *EcoRI* site at the 3' end of the cDNA insert (1,918) is boxed. The putative DNA-binding and ligand-binding domains, designated as regions C and E by analogy to other receptors (see Figs 3 and 5a), are shown in boxes.

Methods. The cDNA libraries used were derived from either MCF-7 or T47D human breast cancer cell lines. Both libraries were randomly primed from total poly(A)⁺ RNA²⁵. *EcoRI* linkers were attached to the cDNAs which were inserted into the *EcoRI* site of the bacteriophage λ vectors gt10 (MCF-7) or gt11 (T47D)²⁵. Recombinant phages from each library (~500,000) were transferred to 150-mm nitrocellulose filters in duplicate and screened using a set of [³²P] end-labelled oligonucleotide probes corresponding to a 24-nt consensus sequence derived from region C sequences of six nuclear receptors (Fig. 1). Hybridization to the probes was carried out at 65°C for 16 h in 6×NET (0.9 M NaCl, 6 mM EDTA, 90 mM Tris-HCl, pH 7.5), 5×Denhardt's solution lacking serum albumin (0.1% Ficoll, 0.1% polyvinyl pyrrolidone) and 0.5% NP40 containing 1 mg ml⁻¹ denatured salmon sperm DNA. The specific activity of the end-labelled oligonucleotide set was >2×10⁸ c.p.m. µg⁻¹. Filters were washed twice for 15 min at room temperature in 4×NET containing 0.5% Denhardt's solution minus albumin, dried and exposed to Kodak XAR-5 film with an intensifying screen at -80°C for 24 h. Positive plaques were re-screened with the oligonucleotide set. Phage DNA was prepared from positive clones and hybridized to nick-translated probes specific for cDNAs of human ER, GR and PR (data not shown). DNA from clones that did not hybridize to these probes were digested with *EcoRI* and the cDNA inserts were re-cloned into pEMBL¹⁶ or Genescribe-Z vectors (USB) for restriction enzyme mapping and DNA sequencing. Step-wise deletions of selected cDNA inserts cloned into pEMBL vectors in both orientations were carried out by digestion with DNase I (ref. 57) to facilitate sequencing. Single-stranded DNA was prepared and sequenced by the chain termination method. GC-rich regions were also sequenced by the chemical method. The sequence of the p63 cDNA insert was confirmed by sequencing a shorter MCF-7 insert and three overlapping T47D cDNAs (see text).

receptors, suggesting that the putative ligand binding to region E of hRAR could indeed be non-steroidal.

High-affinity binding

To identify the ligand which binds to the hRAR protein, the *KpnI-EcoRI* cDNA fragment (Fig. 2) was inserted into the eukaryotic expression vector pSG1 (a gift of S. Green; derived from the vector pKCR2 (ref. 25)), under the control of the SV40 early promoter. The resulting recombinant, RARO, was introduced into HeLa cells grown in the presence of serum treated with dextran-coated charcoal. HeLa cells were also transfected with the parent vector pSG1 in control experiments. HeLa cell cytosol was prepared 40 h after transfection and aliquots were incubated with a variety of radiolabelled prospective ligands at concentrations of 10–20 nM, with and without a 230-fold excess of unlabelled compound. After transfection with RARO, an increase in specific binding of RA was observed (Fig. 4, compare transfections with RARO and pSG1, and data not shown). This increase in RA binding following transfection with RARO was not inhibited by a 230-fold excess of testosterone, vitamin D3, triiodothyronine (T3) or reverse T3 (rT3), whereas some competition was achieved with retinol or retinal. At lower concentrations, however, retinol and retinal were at least 10–20 times less efficient than RA at competing for this binding. On the other hand, RA binding to endogenous HeLa cells (pSG1 control transfection) could not be inhibited efficiently with either retinol or retinal, but could be with RA (Fig. 4). In this respect we note that the binding of labelled RA by the cellular retinoic acid-binding protein CRABP, although inhibited by RA, is not sig-

nificantly inhibited by retinol (see refs 2, 6 and 19). In other experiments (not shown) we found that the increase in specific binding of labelled RA to the cytosol of HeLa cells transfected with RARO was maximal at concentrations of ~20 nM and reduced by at least 10-fold at 1 nM. As CRABP is present in HeLa cells, it is not possible to determine an accurate *K_d* value for RA binding to hRAR, but these results indicate that it should be lower than 10 nM.

Thus, the hRAR protein appears to bind RA selectively with high affinity, and bind retinol and retinal with lower affinity. The hRAR protein is therefore a member of the nuclear receptor family which appears to be a receptor for RA.

Retinoic acid-dependent *trans*-activation

By analogy with other members of the nuclear receptor family, it was likely that the hRAR could act as an inducible enhancer-binding factor. This hypothesis would be best tested by transfecting HeLa cells with the hRAR expression vector RARO together with a reporter recombinant containing the RA-responsive enhancer element of a target gene whose transcription is controlled by RA. Although such genes may have been identified (see above for refs) their promoter responsive elements are not yet well characterized. But it has been shown previously that replacing region C of the hER with region C of hGR results in a chimaeric receptor 'ER-GR.CAS' which can activate the glucocorticoid-responsive reporter gene *MMTV-CAT* in the presence of oestradiol^{31,32}. Also, replacing region C of hGR with region C of hER results in a chimaera 'GR-ER.CAS' which can activate the oestrogen-responsive reporter gene *vit-tk-CAT* in

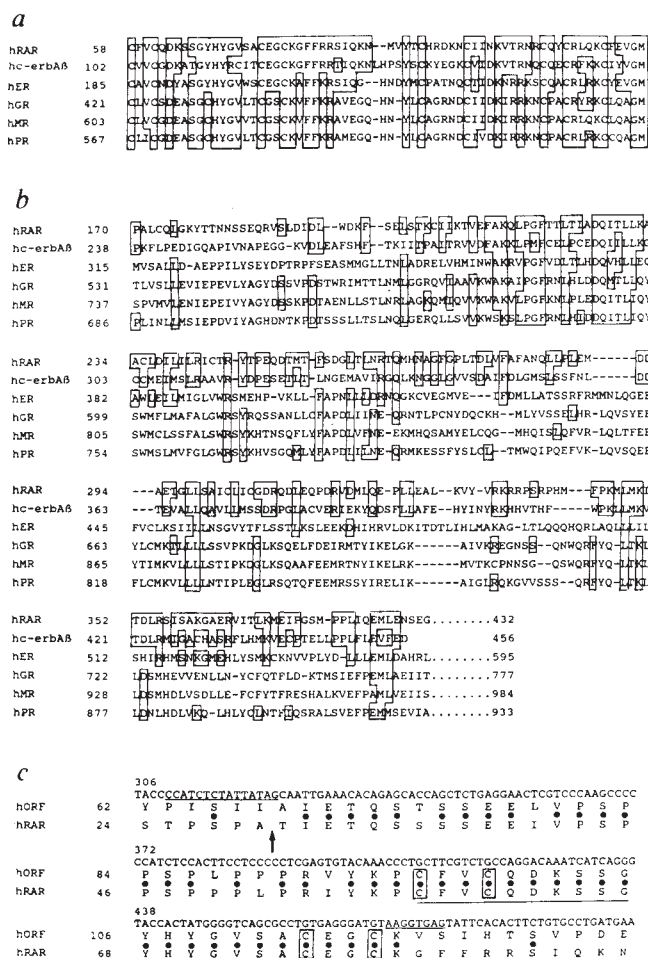


Fig. 3 Homology between the predicted sequences of hRAR and human nuclear receptors. *a*, The sequence of the putative DNA-binding domain (region C in Fig. 2) of hRAR is compared with those of the human nuclear receptors for thyroid hormone (h α erb A β)²⁷, oestrogen (hER)²⁵, glucocorticoid (hGR)⁴⁰, mineralocorticoid (hMR)³⁰ and progesterone (hPR)⁴¹. *b*, A similar alignment for region E of hRAR (Fig. 2) and the corresponding ligand-binding regions E of the same receptors. Dashes, spaces introduced into the sequences to obtain maximal alignment. Amino acid residues conserved between hRAR and at least one other receptor in these comparisons are boxed. *c*, Nucleotide and predicted amino acid sequences of the open reading frame for a novel human nuclear receptor-like genomic sequence (hORF), isolated from a hepatocellular carcinoma⁴⁴, aligned with the corresponding sequence of hRAR. The hORF nucleotide positions are numbered above the DNA sequence⁴⁴. Dots, amino acid residues conserved between the two sequences. Proposed acceptor and donor splice sites for the hORF DNA sequence are underlined at the top left and the bottom right of the alignment, respectively. The portion of hRAR sequence corresponding to region C is underlined. The conserved cysteine residues are boxed. Arrow, see text. Amino acid positions are numbered on the left-hand side.

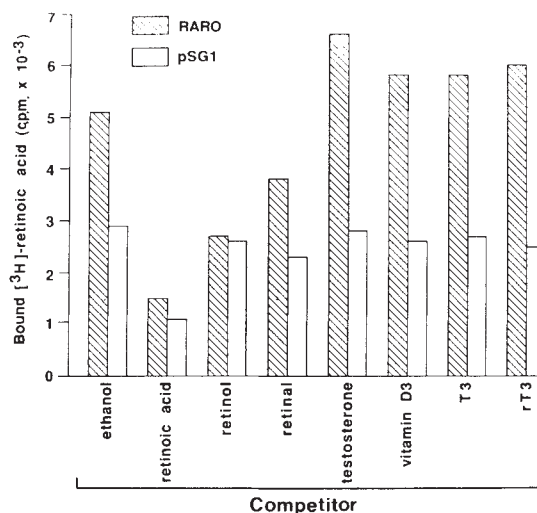


Fig. 4 Specific binding of RA to hRAR protein synthesized in HeLa cells transfected with the hRAR cDNA expression vector RARO. While cell lysates from HeLa cells transfected with either the *in vivo* expression vector RARO (see legend to Fig. 5) containing the hRAR cDNA open reading frame (hatched bars) or the parent vector pSG1 (open bars) were incubated with 17 nM [³H]RA in the presence of vehicle (ethanol) or a 230-fold excess of the unlabelled competitors as indicated. The amount of bound [³H] RA is shown.

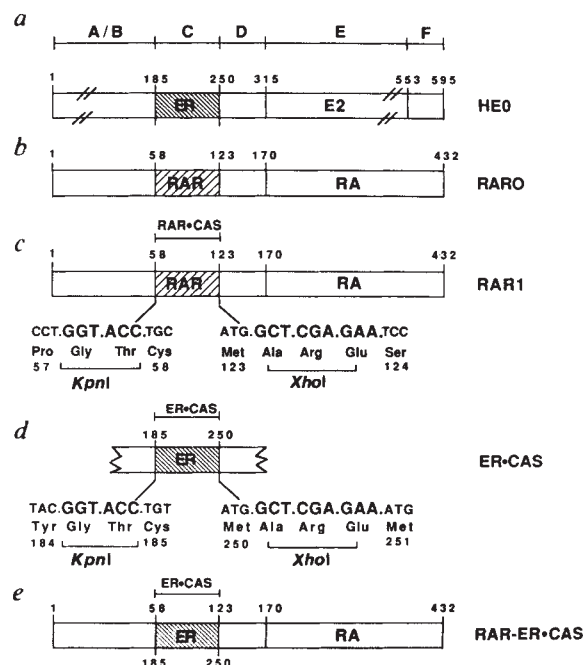
Methods. HeLa cells ($\sim 10^6$ per dish) were transfected as previously described³⁵ with 20 μg of either the parent vector pSG1, an *in vivo* expression vector analogous to pKCR2²⁵ that contains the SV40 early promoter linked to a rabbit β -globin intervening sequence, or RARO. After 40 h, the cells were harvested, collected by centrifugation at low speed, washed with cold phosphate-buffered saline and resuspended in 120 μl (per dish) of 20 mM Tris-HCl, pH 7.2, 1 mM EDTA, 2 mM dithiothreitol (DTT), 50 mM NaCl and 0.3 mM phenylmethane sulphonyl fluoride (PMSF). The suspension was frozen and thawed once, and the cells were pelleted by centrifugation at 10,000 g for 15 min. All manipulations were carried out at 4°C. The supernatants from several dishes were pooled and the protein concentration was determined by Bradford assay. For the binding-competition assays, 1 μl of [³H]RA (New England Nuclear, 52.5 Ci mmol⁻¹) was added to 50 μl of lysate, giving a final concentration of 17 nM. Competing ligand (2 μl) was added to a final concentration of 4.2 μM . The competitors tested were all *trans*-retinoic acid, all *trans*-retinol, all *trans*-retinal, testosterone, L-3,5,3'-triiodothyronine (T3), L-3,3',5'-triiodothyronine (rT3) (Sigma) or 1,25-dihydroxyvitamin D3 (Hoffman-La Roche). The lysates were incubated for 3 h at 4°C and mixed with 100 μl of a cold dextran-coated charcoal suspension for 15 min. Unbound [³H]RA was removed by centrifugation at 10,000 g for 3 min at 4°C and 100 μl of the supernatant was taken for scintillation counting. Values shown are from duplicate experiments and were corrected for differences in protein concentration between the pSG1-transfected and RARO-transfected pooled lysates.

logen A2 gene (*vit*) located upstream of the herpes virus thymidine kinase promoter (*tk*) and the *Escherichia coli* chloramphenicol acetyl transferase (*CAT*) gene³¹.

Cotransfection of HeLa cells with the chimaeric receptor RAR-ER.CAS and the *vit-tk-CAT* reporter gene resulted in an increase of CAT activity which was dependent on the presence of 10^{-8} M RA, but not of oestradiol (Fig. 6a, lanes 7 and 8, 9–11, and 12 and 13; three different transfection experiments). Using 1 μ g of *vit-tk-CAT* construct, maximal stimulation was achieved with ~ 100 ng of RAR-ER.CAS (Fig. 6b). We note that the stimulation of *vit-tk-CAT* expression by the RAR-ER.CAS chimaeric receptor in the presence of RA is comparable to that brought about under similar conditions by the hER expression vector HEO in the presence of oestradiol (Fig. 6a, lanes 12–15; HeLa cells transfected under these conditions contained on

Fig. 5 Construction of the chimaeric receptor RAR-ER.CAS. The receptor expression vectors HEO, RARO, RAR1, ER.CAS and RAR-ER.CAS are shown schematically in the figure. Numbers designate amino acid positions. The location of regions A-F previously defined for the human (HEO) and chicken oestrogen receptors²⁶ are indicated at the top of the figure. The putative DNA-binding domain C of HEO spans 66 consecutive amino acid residues (185–250). RARO contains an homologous region (see Figs 2, 3 and text) also spanning 66 amino acid residues (58–124). Unique restriction sites were created flanking the 5' (*Kpn*I) and 3' (*Xho*I) ends of region of RARO (to create RAR1 which contains the excisable RAR-CAS cassette). RAR-ER.CAS was then constructed by replacing the RAR.CAS cassette in RAR1 with the corresponding region C cassette ER.CAS present in HE28³¹.

Methods. The 1.7-kb *Kpn*I–*Eco*RI fragment of the hRAR cDNA was excised and an *Eco*RI–*Kpn*I linker was attached to destroy the *Kpn*I recognition sequence. The fragment was ligated into the *Eco*RI site of the expression vector pSG1 which was derived from pKCR2 (ref. 25) giving the hRAR expression vector RARO. Unique *Kpn*I and *Xho*I restriction enzyme sites were introduced into RARO in a similar way to that described for HE28³¹, except that the cDNA encoding hRAR was inserted into Genescribe-Z pTZ19U for the oligonucleotide-directed mutagenesis steps. The resulting recombinants RAR1 and RAR.ER.CAS were then inserted into the eukaryotic expression vector pSG1.



average ~1,000 molecules per cell of either RAR-ER.CAS or hER receptors as determined by binding of labelled RA or oestradiol, respectively). No such stimulation of *vit-tk-CAT* expression was observed after transfection with either the parent vector pSG1, RARO (which expresses the hRAR protein) or RAR1, which contains the restriction enzyme sites used to generate the RAR-ER.CAS receptor (Fig. 6a, lanes 1–6; see also Fig. 4); we have no explanation for the 2–3-fold stimulation of CAT activity when RA was present in these cotransfection experiments). Activation of *vit-tk-CAT* expression by the RAR-ER.CAS receptor was specific for this ERE-containing reporter gene, as no stimulation was seen using the glucocorticoid-responsive reporter gene *MMTV-CAT*, which was activated by the GR expression vector HG1 in the presence of dexamethasone (Fig. 6a, lanes 16–19). That the stimulation of *vit-tk-CAT* expression was specific for RA is shown in Fig. 6c where various compounds (concentration 10^{-8} M) were tested for their ability to activate *vit-tk-CAT* expression in the presence of the chimaeric receptor. Maximal activation of *vit-tk-CAT* expression was achieved between 10^{-8} and 10^{-7} M RA, but stimulation was already observed at 10^{-9} M (Fig. 6d, lanes 1–7; compare with the parent vector pSG1, lanes 13–18). We note that this optimal RA concentration for activation of *vit-tk-CAT* expression is similar to the optimal concentration of RA for increased RA binding in the cytoplasm of RARO-transfected HeLa cells (see above). A stimulation of *vit-tk-CAT* expression could also be observed in the presence of retinol, after cotransfection with RAR-ER.CAS (Fig. 6d, lanes 8–12). Retinol was at least 100-fold less effective than RA, however (Fig. 6d, compare lanes 4 and 5 with lanes 11 and 12). Essentially the same results (data not shown) were obtained when *trans*-activation by RAR-ER.CAS was determined by quantitative S1 nuclease analysis of RNA initiated at the *tk* promoter start site using *vit-tk*-globin as the reporter gene, in which the CAT sequence of *vit-tk-CAT* has been replaced by the rabbit β -globin sequence from position –9 to 1,650 (ref. 32).

Related human genomic sequence

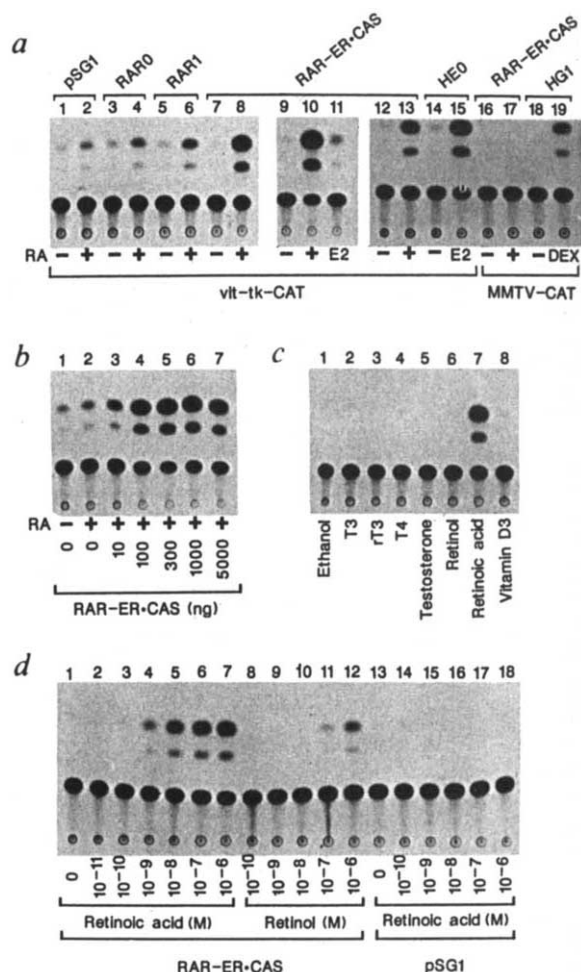
There is a remarkable identity between the N-terminal half of region C of hRAR and an amino-acid sequence encoded in a DNA segment located immediately downstream from the insertion site of the hepatitis B virus (HBV) genome in a human hepatoma⁴⁴. As pointed out by Tiollais and his collaborators⁴⁴,

this region of the human genome contains an open reading frame (hORF) encoding an amino-acid sequence that is highly similar to the N-terminal half of the putative DNA-binding domain of nuclear receptors (see Fig. 3c; region C of hRAR is underlined). In particular, this segment of hORF encodes the two pairs of cysteines (boxed in Fig. 3c) proposed to be essential for the folding of this region of the nuclear receptors into a DNA-binding finger^{34,45}. It is striking that the deduced amino-acid sequence of this putative finger of hORF is identical to that of hRAR, and that this homology extends further upstream up to amino acid 31 of hRAR, but ceases abruptly after amino acid 79, where there is a putative splicing donor site (underlined in Fig. 3c) in the hORF genomic DNA⁴⁴. Interestingly, amino acids 29 and 30 of hRAR are encoded in two separate exons of the hRAR gene (our unpublished data; arrow in Fig. 3c) and there is a putative splicing acceptor site (underlined in Fig. 3c) at the corresponding position of the hORF genomic DNA⁴⁴. Note that the two putative DNA-binding fingers that are present in region C of PR⁴⁵, ER (unpublished results from our laboratory) and the *c-erbA* product⁴⁶ are encoded in two separate exons.

The above observations suggest very strongly that the gene encoding hRAR and the putative gene containing hORF have evolved by duplication of an ancestral gene. That the two genes are clearly different, however, is indicated by comparison of the nucleotide sequences shown in Figs 2 and 3c, which reveals numerous differences, and also by the observation of Tiollais's group that the hORF sequence maps to human chromosome 3 (ref. 44), whereas we have found that the hRAR gene is localized on chromosome 17q21.1 (ref. 47). Two *v-erbA*-related genes, *erbA1* and *erbA2*, have been previously mapped to chromosome 17. *erbA1*, which maps to 17q11.2→q12, probably corresponds to the thyroid hormone receptor hTR α (see refs 28, 48). *erbA2* has approximately the same chromosomal localization (17q21.3) as hRAR⁴⁸. It is however different from hRAR, which does not hybridize to the specific *erbA2* probes of Jansson *et al.*⁴⁹ (unpublished data). In addition, another member of the *erbA2* subgroup appears to be located at 17q25⁴⁸. In view of the similarities pointed out above between regions E of the hRAR and *erbA* genes, it seems likely that the gene for hRAR and all three *erbA* genes located on chromosome 17, as well as the thyroid-hormone receptor hc-*erbA* β (htRK β , ref. 27) and hORF⁴⁴ genes, which are both located on chromosome 3, have evolved by duplications

Fig. 6 *a*, RAR-ER.CAS activates transcription of the *vit-tk-CAT* reporter plasmid, but not *MMTV-CAT*, in the presence of RA. HeLa cells were transfected with the expression vectors pSG1 (lanes 1 and 2), RARO (lanes 3 and 4), RAR1 (lanes 5 and 6), RAR-ER.CAS (lanes 7–13, 16 and 17), HEO (lanes 14 and 15) or HG1 (lanes 18 and 19) (an hGR expression vector, see ref. 32) with either the oestrogen-responsive *vit-tk-CAT*³¹ (lanes 1–15) or glucocorticoid-responsive *MMTV-CAT*³¹ (lanes 16–19) reporter plasmids (see text). Cells were treated with vehicle (ethanol) denoted by minus (–), 10^{-8} M retinoic acid (RA+), 10^{-8} M oestradiol (E2) or 10^{-7} M dexamethasone (DEX) as indicated. *b*, Activation of CAT expression by RA is dependent upon the amount of transfected RAR-ER.CAS. HeLa cells were co-transfected with *vit-tk-CAT* and the indicated range of amounts of RAR-ER.CAS, and grown in the presence of 10^{-8} M RA (lanes 2–7) or vehicle (lane 1). *c*, Retinoic acid is a specific inducer of trans-activation by RAR-ER.CAS chimaeric receptor. RAR-ER.CAS was co-transfected into HeLa cells with *vit-tk-CAT* in the presence of 10^{-8} M of the metabolites indicated in the figure. *d*, Comparison of RA and retinol as inducers of trans-activation by the RAR-ER.CAS chimaeric receptor. HeLa cells were transfected with either RAR-ER.CAS (lanes 1–12) or pSG1 (lanes 13–18), together with the *vit-tk-CAT* reporter plasmid. The cells were treated with increasing concentrations of either RA (lanes 1–7 and 13–18) or retinol (lanes 8–12) as shown in the figure.

Methods. Transfection experiments were carried out as follows unless otherwise stated below. HeLa cells ($\sim 10^6$ per dish) were transfected using calcium phosphate³⁵ with 1 μ g of the expression vector pSG1, RARO, RAR1, RAR-ER.CAS or HG1 (except for lanes 12–15, panel *a*, where 100 ng of expression vectors RAR-ER.CAS or HEO were used), along with 1 μ g of the reporter plasmids *vit-tk-CAT* or *MMTV-CAT* and 3 μ g of pCH110 (Pharmacia), a β -galactosidase expression vector used as an internal control to normalise for transfection efficiency. The total amount of DNA transfected was made up to 20 μ g with carrier DNA (Bluescribe M18+). After a 20 h transfection period, cells were washed with serum-free medium, fed with medium containing charcoal-treated serum and incubated for an additional 24 h with the compounds indicated. Cells were harvested and cytosolic extracts prepared by subjecting the cells to three consecutive freeze-thaw cycles followed by centrifugation at 10,000 g for 15 min. Aliquots were normalized for β -galactosidase activity. Typically, aliquots corresponding to 15 units of β -galactosidase activity for each sample were assayed for CAT activity using 0.1 μ Ci of [¹⁴C]-chloramphenicol (Amersham). In *c*, HeLa cells were transfected with 5 μ g of RAR-ER.CAS along with 1 μ g of the reporter plasmid *vit-tk-CAT* and 14 μ g of carrier DNA. L-3,5,3',5'-triiodothyronine (T3), L-3,5,3',5'-triiodothyronine (rT3), L-3,5,3',5'-tetraiodothyronine (T4), testosterone, retinol, retinoic acid (Sigma) or 1,25-dihydroxyvitamin D3 (Hoffman-La Roche) were added at 10^{-8} M as indicated. In *d*, HeLa cells were transfected with 100 ng of the expression vectors RAR-ER.CAS or pSG1, 400 ng of *vit-tk-CAT* and carrier DNA to 20 μ g.



from a common progenitor which itself diverged early in evolution from the steroid hormone receptor ancestor.

Discussion

We have cloned a cDNA corresponding to a new member of the nuclear receptor multigene family. The encoded protein specifically binds the vitamin A derivative, retinoic acid, with high affinity and has the properties expected for an inducible enhancer factor functioning like a steroid hormone receptor. Trans-activation was detectable with RA concentrations as low as 10^{-9} M and was optimal at concentrations higher than 10^{-8} M. In contrast, concentrations of retinol at least 100-fold higher were required to achieve the same effect. These concentrations are those at which RA and retinol promote differentiation of F9 embryonal carcinoma stem cells and induce the transcription of specific genes in these cells^{14,16,50}. They also correspond to the RA and retinol concentrations which inhibit the ability of malignant cells to form colonies under anchorage-dependent growth conditions^{37,38}. Moreover, the tissue concentrations at which RA appears to act as a morphogen in the developing chick limb bud are also in the same nanomolar range (see refs 10 and 12). These correlations suggest very strongly that the retinoic acid receptor (hRAR) that we have cloned is the transcriptional mediator of at least some of these fundamental biological processes. Although the *in vitro* competition experiments shown in Fig. 4 indicate that high concentrations of retinol can compete

for RA binding to hRAR, the *in vivo* trans-activation induced by high concentrations of retinol (Fig. 6) may be due to conversion of retinol to RA, a process known to occur in cultured cells¹⁴.

Using hRAR cDNA as a probe, Northern blotting analysis of poly(A)⁺ RNA isolated from MCF-7 and T47D cells has revealed the existence of two rare and equally represented 3.5- and 5.1-kb messenger RNA species, in amounts not very different from those of hER mRNA in the same cells (data not shown). As the gene for hRAR appears to be unique (unpublished data), these two RNA species probably correspond to alternative processing products of the same transcription unit. They appear to be conserved during evolution and to be widely distributed, as they are present, albeit in even lower amounts, in poly(A)⁺ RNA of human HeLa cells and a variety of mouse tissues (for example, brain, thymus, spleen, testis, lung, liver, prostate and kidney—our unpublished results). Why then has hRAR not been previously detected? As noted in the introduction, a 15.5K cellular protein, CRABP, which specifically binds RA but not retinol, is widely distributed in a variety of tissues and cells in culture where it appears to be rather abundant (several tens to hundreds of thousands molecules per cell)^{6,9,19}. If the hRAR protein, like steroid hormone receptors, is present in only a few thousand molecules per cell, as suggested by the low abundance of its mRNA (see above), then the presence of CRABP may have precluded its detection. Note in this respect that the increase in RA binding which could be inhibited with high concentrations

of retinol after transfection with RAR α corresponds in Fig. 4, on average, to the presence of at least 60,000 hRAR molecules per cell.

It has been previously proposed that CRABP could function as a shuttle protein delivering RA to undefined nuclear acceptor sites themselves unable to bind free RA (see introduction). It is unlikely that hRAR could simply correspond to the RA nuclear acceptor sites, as these appear to be present in a high copy number (several tens to hundreds of thousands per nucleus; see refs 18, 23). Whether CRABP and hRAR cooperate in the mechanism of action of RA remains to be investigated. We note that some, but not all, mouse mutant embryonal carcinoma cells that fail to differentiate in response to RA have undetectable levels of CRABP (see refs 51, 52). Similarly some, but not all, tumour cells selected for RA resistance have undetectable CRABP levels^{38,53}.

The nature of the putative protein encoded in the hORF gene⁴⁴ is unknown. It is striking that what appears to be the N-terminal half of the hORF region C exhibits a 100% amino acid identity with the corresponding region of hRAR (Fig. 3c). Should the C-terminal halves also be identical, then both proteins may activate transcription of the same genes, as this region of the nuclear receptors is known to be responsible for the selectivity of transcriptional activation of the target genes^{31,32}. Assuming that the ligand of the putative hORF protein is also RA, the situation would then be analogous to that for thyroid hormones, where at least two cell-specific receptors appear to bind the same hormones²⁸. Alternatively, the putative hORF protein may bind a different ligand, possibly retinol because a cellular retinol-binding protein exists in many cells and tissues (see above). Note in this respect that retinol is important for the maintenance of some epithelial cells, like the germinal epithelial cells of the testis which degenerate in the absence of retinol, even when RA is present (see ref. 7).

The approach used here to clone hRAR could obviously be

used to isolate further members of the nuclear receptor multigene family which may be expressed in embryonic or adult tissues. Chimaeric receptors similar to RAR-ER.CAS should be useful to identify unknown natural ligands or putative ligands unavailable in labelled forms. For instance, the addition of tissue extracts to the growth medium of HeLa cells cotransfected with the reporter *vit-tk-CAT* and a chimaeric receptor containing hER region C and the ligand-binding domain of the new member of the family, could be used to detect the presence of the cognate ligand. Finally, based on the case for thyroid hormone receptor *v-erbA*, we speculated that the nuclear receptor multigene family may contain several potential oncogenes¹⁷. The finding⁴⁴ that, in a human hepatoma, the hepatitis B virus genome is inserted into a genomic sequence (hORF) closely related to that of the hRAR gene raises the possibility that altered retinoid receptors may have oncogenic properties. Such a possibility would not be too surprising in view of the crucial role played by the vitamin A derivatives in embryogenesis, differentiation and tumour cell growth.

We thank the hormone receptor group for interesting discussions, in particular S. Green for reading the manuscript and providing HEO, ER.CAS, pSG1 and the T47D cDNA library, and V. Kumar and M. Govindan for providing HG1. The 1,25-dihydroxyvitamin D3 was generously provided by Hoffmann-La Roche. We also thank A. Staub and F. Ruffenach for synthesizing the oligonucleotides, the tissue culture lab for medium and cells, C. Werlé and B. Boulay for preparing figures and the secretariat for typing the manuscript. Supported by INSERM (grants CNAMTS), CNRS (AIV), the Ministère de la Recherche et de l'Enseignement Supérieur, the Fondation pour la Recherche Médicale and the Association pour la Recherche sur le Cancer. M.P. is the recipient of a post-doctoral fellowship from the Medical Research Council of Canada, and N.J.B. was the recipient of a Royal Society post-doctoral fellowship.

Received 26 October; accepted 10 November 1987.

1. Lotan, R. *Biochim. Biophys. Acta* **605**, 33-91 (1980).
2. Chytil, F. *Pharmacol. Rev.* **36**, 93S-100S (1984).
3. Roberts, A. B. & Sporn, M. B. in *The Retinoids* (eds Sporn, M. B., Roberts, A. B. & Goodman, D. S.) 209-286 (Academic, Florida, 1984).
4. Strickland, S. *et al. Cancer Res.* **43**, 5268-5272 (1983).
5. Shubaita, H. E., Sambrook, J. F. & McCormick, A. M. *Proc. natn. Acad. Sci. U.S.A.* **84**, 5645-5649 (1987).
6. Haussler, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 5525-5529 (1983).
7. Porter, S. B., Ong, D. E., Chytil, F. & Orgebin-Crist, M.-C. *J. Androl.* **6**, 197-212 (1985).
8. Maden, M. *Trends Genet.* **1**, 103-107 (1985).
9. Maden, M. & Summerbell, D. J. *Embryol. exp. Morph.* **97**, 239-250 (1986).
10. Thaller, C. & Eichele, G. *Nature* **327**, 625-628 (1987).
11. Wilde, S. M., Wedden, S. E. & Tickle, C. *Development* **100**, 723-733 (1987).
12. Slack, J. M. W. *Nature* **327**, 553-554 (1987).
13. Kurkinen, M., Barlow, D. P., Helfman, D. M., Williams, J. G. & Hogan, B. L. M. *Nucleic Acids Res.* **11**, 6199-6209 (1983).
14. Williams, J. B. & Napoli, J. L. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4658-4662 (1985).
15. Wang, S.-Y., LaRosa, G. J. & Gudas, L. J. *Dev. Biol.* **107**, 75-86 (1985).
16. Alonso, A., Weber, T. & Jorcano, J. L. *Roux's Arch. Dev. Biol.* **196**, 16-21 (1987).
17. Green, S. & Chambon, P. *Nature* **324**, 615-617 (1986).
18. Barkai, U. & Sherman, M. I. *J. Cell Biol.* **104**, 671-678 (1987).
19. Ong, D. E., Crow, J. A. & Chytil, F. *J. biol. Chem.* **257**, 13385-13389 (1982).
20. Grippo, J. F. & Gudas, L. J. *J. biol. Chem.* **262**, 4492-4500 (1987).
21. Kato, M., Kato, K., Blaner, W. S., Chertow, B. S. & Goodman, D. S. *Proc. natn. Acad. Sci. U.S.A.* **82**, 2488-2492 (1985).
22. Wei, L. N., Mertz, J. R., Goodman, D. S. & Nguyen-Hu, M. C. *Molec. Endocrin.* **1**, 526-534 (1987).
23. Takase, S., Ong, D. E. & Chytil, F. *Arch. Biochem. Biophys.* **247**, 328-334 (1986).
24. Weinberger, C., Hollenberg, S. M., Rosenfeld, M. G. & Evans, R. M. *Nature* **318**, 670-672 (1985).
25. Green, S. *et al. Nature* **320**, 134-139 (1986).
26. Krust, A. *et al. EMBO J.* **5**, 891-897 (1986).
27. Weinberger, C. *et al. Nature* **324**, 641-646 (1986).
28. Thompson, C. C., Weinberger, C., Lebo, R. & Evans, R. M. *Science* **237**, 1610-1614 (1987).
29. McDonnell, D. P., Mangelsdorf, D. J., Pike, J. W., Haussler, M. R. & O'Malley, B. W. *Science* **235**, 1214-1217 (1987).
30. Arriza, J. L. *et al. Science* **237**, 268-275 (1987).
31. Green, S. & Chambon, P. *Nature* **325**, 75-78 (1987).
32. Kumar, V. *et al. Cell* (in the press).
33. Enver, T. *Nature* **317**, 385-386 (1985).
34. Green, S., Gronemeyer, H. & Chambon, P. in *Growth Factors and Oncogenes in Breast Cancer* (ed. Sluyser, M.) 7-28 (Horwood, Chichester, 1987).
35. Kumar, V., Green, S., Staub, A. & Chambon, P. *EMBO J.* **5**, 2231-2236 (1986).
36. Gronemeyer, H. *et al. EMBO J.* (in the press).
37. Lacroix, A. & Lippman, M. E. *J. clin. Invest.* **65**, 586-591 (1980).
38. Ueda, H., Ono, M., Hagino, Y. & Kuwano, M. *Cancer Res.* **45**, 3332-3338 (1985).
39. Govindan, M. V., Devic, M., Green, S., Gronemeyer, H. & Chambon, P. *Nucleic Acids Res.* **13**, 8293-8304 (1985).
40. Hollenberg, S. M. *et al. Nature* **318**, 635-641 (1985).
41. Misrahi, M. *et al. Biochem. biophys. Res. Comm.* **143**, 740-748 (1987).
42. Hollenberg, S. M., Giguere, V., Segui, P. & Evans, R. M. *Cell* **49**, 39-46 (1987).
43. Rusconi, S. & Yamamoto, K. R. *EMBO J.* **6**, 1309-1315 (1987).
44. Dejean, A., Bougueleret, L., Grzeschik, K. H. & Tiollais, P. *Nature* **322**, 70-72 (1986).
45. Jeltsch, J. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 5424-5428 (1986).
46. Zahraoui, A. & Cuny, G. *Eur. J. Biochem.* **166**, 63-69 (1987).
47. Mattei, M. G., Mattei, J. F., Petkovich, M. & Chambon, P. *Hum. Genet.* (submitted).
48. Gosden, J. R., Middleton, P. G., Rout, D. & De Angelis, C. *Cytogenet. Cell Genet.* **43**, 150-153 (1986).
49. Jansson, M., Philipson, L. & Vennstrom, B. *EMBO J.* **2**, 561-565 (1983).
50. Jettten, A. & De Luca, L. M. *Biochem. biophys. Res. Comm.* **114**, 593-599 (1983).
51. McCue, P. A., Matthaei, K. I., Taketo, M. & Sherman, M. I. *Dev. Biol.* **96**, 416-426 (1983).
52. Wang, S.-Y. & Gudas, L. J. *J. biol. Chem.* **259**, 5899-5906 (1984).
53. Lotan, R., Stolarsky, T. & Lotan, D. *Cancer Res.* **43**, 2868-2875 (1983).
54. Debuire, B. *et al. Science* **224**, 1456-1459 (1984).
55. Miesfeld, R. *et al. Cell* **46**, 389-399 (1986).
56. Dente, L., Cesareni, G. & Cortese, R. *Nucleic Acids Res.* **11**, 1645-1655 (1983).
57. Lin, H.-C., Lei, S.-P. & Wilcox, G. *Analyt. Biochem.* **147**, 114-119 (1985).

Galaxy distribution in a cold dark matter universe

Simon D. M. White*, Marc Davis†, George Efstathiou‡
& Carlos S. Frenk§

* Steward Observatory, University of Arizona, Tucson,
Arizona 85721, USA

† Departments of Astronomy and Physics, University of California,
Berkeley, California 94720, USA

‡ Institute of Astronomy, Madingley Road,
Cambridge CB3 0HA, UK

§ Physics Department, University of Durham, South Road,
Durham DH1 3LE, UK

The observed galaxy distribution is almost our sole source of information about the distribution of matter on very large scales. Most efforts to estimate the mean density of the Universe from the dynamical behaviour of observed inhomogeneities assume the galaxy distribution to trace that of the mass¹⁻⁴. If galaxies are assumed to be a statistically fair sample from the overall mass distribution, the Universe must be open, but if galaxies are systematically overrepresented in high density regions then the observational data may be compatible with a closed or flat universe^{5,6}. There are many reasons to suspect that the galaxy distribution may indeed be biased in this way with respect to the mass⁷⁻¹⁰, and, with a particular model for the biasing, a flat universe dominated by cold dark matter (CDM) provides a remarkably good description of observed structure^{6,11,12}. Here we show that the gravitational growth of structure by hierarchical clustering leads automatically to a bias in the distribution of galaxies. The strength of this bias is such that if the Universe does indeed conform to the CDM model then its density must approach the closure value. The mechanism predicts a significant dependence of the strength of galaxy clustering on the depth of the potential well of the galaxies considered.

The hierarchical clustering model assumes structure to evolve by the continual inhomogeneous aggregation of matter into larger and larger units¹³. The apparently hierarchical distribution of galaxies reflects the present scale of structure formation, and galaxies themselves are supposed to have formed at an earlier stage of this same process. The high density-contrast of the luminous parts of galaxies and the clear spatial separation between stars and dark matter suggest that the visible material contracted dissipatively within a virialized dark halo¹⁴. The characteristic luminosity of galaxies then follows from the characteristic cooling time of the protogalactic gas¹⁴⁻¹⁶, and the angular momentum of spirals agrees well with predictions from the tidal torque theory¹⁷. The most plausible origin for ellipticals in this picture may be the merging of pre-existing stellar or partially stellar systems^{18,19}. The CDM theory predicts a specific set of initial conditions from which this kind of evolution can proceed^{5,20}. Its success lies in the fact that these initial conditions seem to be what is required to obtain the correct abundance and structure of systems ranging from galaxy haloes to rich galaxy clusters^{6,12,21,22} (but see below). Within such models the distribution of galaxies reflects the distribution of the centres of all haloes which ever survived long enough for gas to cool¹⁴. This distribution is modified by mergers and by other physical processes after galaxy formation. To evaluate possible biases we must then compare the statistical properties of this distribution to those of the mass as a whole.

We have attempted to use direct numerical simulation to relate the galaxy and mass distributions within the CDM theory. This is difficult because a successful model must have sufficient resolution to detect the formation of a modest galaxy and yet be large enough to allow a reliable measurement of galaxy clustering. We decided to simulate cubes of side $2,500 \text{ km s}^{-1}$

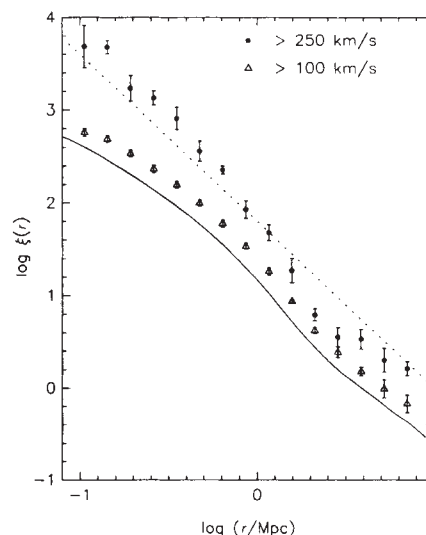


Fig. 1 Spatial autocorrelation function $\xi(r)$ for 'galaxies', as defined in the text, and for the mass distribution in the simulations. The solid line shows the mass correlations, and open triangles and filled circles show the correlation function for 'galaxies' with $V_c > 100$ and 250 km s^{-1} , respectively. The error bars show the error in the mean derived from the three simulations. The dashed line shows the power-law $\xi(r) = (r/10 \text{ Mpc})^{-1.8}$, which approximately describes the clustering of bright galaxies¹.

using 262,144 particles of mass $3.5 \times 10^{10} M_\odot$ (here and below we assume Hubble's constant $H_0 = 50 \text{ km s}^{-1} \text{ Mpc}^{-1}$ and density ratio to the critical value $\Omega = 1$). Our force resolution is about $50/(1+z) \text{ kpc}$, where z is redshift. We generate random phase initial conditions using Zeldovich's formulation of linear theory, as described previously²⁴. Using the P³M technique^{23,24} we can follow such regions from $z = 6$ to the present day in about 20 h of computer time on a Cray XMP. As a result of our compromises we are unable to model the formation of galaxies with characteristic circular velocities smaller than about 100 km s^{-1} , and we are likely to underestimate the strength of clustering on scales larger than about 5 Mpc. We have run three models taking the initial fluctuation amplitude and all other cosmological parameters from our previous work on flat CDM universes^{6,12,21,22}. We thus allow ourselves no freedom to adjust length, mass and timescales when analysing the mass distribution in these simulations, but to study galaxy clustering we must still model galaxy formation within CDM haloes.

In these models, as in previous simulations of hierarchical clustering, we find the substructure of clumps to be erased soon after they collapse to high density. A 'group of galaxies' thus appears as a particularly massive but nonetheless smooth clump of dark matter. If hierarchical clustering of this kind is to mimic the real Universe, galaxies must somehow survive the merging of their haloes in clusters. On the other hand, some mergers between galaxies seem inevitable and may even be desirable to form elliptical galaxies. Our simulations follow the evolution of the collisionless dark matter component only, and so do not allow us to study the behaviour of the dissipative gas from which galaxies must form. We are thus forced to put in galaxy formation and galaxy merging 'by hand' in a way which, although plausible, remains somewhat *ad hoc*. We have adopted the following algorithms.

Each time a model expands by a factor $2^{1/3}$ we list the positions of all the particles and locate high density regions. These we identify as haloes of dark matter. We identify the most strongly bound particle in each halo as the 'galaxy' which has condensed within it. We then find the mass contained in a surrounding sphere of mean density 1,000 times the *present* critical density, and define the associated velocity, $V_c = (GM(r)/r)^{1/2}$, as the circular velocity of the galaxy. The density profiles of haloes are

such that this velocity depends very little on our choice of overdensity^{21,22}. A problem arises because any given halo may have been present at an earlier time, when a different particle is likely to have been identified as its central galaxy. To avoid multiple galaxy formation within the same halo we adopt the following procedure. For each halo we find the distance d from the central particle to every old galaxy within 0.5 Mpc. The old galaxy with the largest value of V_c/d is taken to be the previous central object and is deleted, while the new central galaxy is assigned the maximum of its own V_c and that of its previous incarnation. To simulate merging, we consider any other old galaxy within a distance r_{mer} of the central object to be drawn into it by dynamical friction. Such objects are also deleted and the central galaxy is again assigned the maximum of its own V_c and that of its victim. We adopt $r_{\text{mer}} = 100$ kpc as our standard value but have also tried 0 and 200 kpc. Notice that these procedures assign a galaxy to every halo that ever forms. As a result we end up with a few 'galaxies' with very large circular velocities at the centres of galaxy groups. In the real world, long cooling times are expected to inhibit the formation of such systems.

Applying these algorithms to our three models we find that at the present day they contain 3,806, 564, and 1 dark haloes with circular velocities exceeding 100, 200 and 800 km s^{-1} respectively. This last object is a system with about the mass, velocity dispersion and galaxy content of the Virgo cluster²⁵. Within these haloes we find 7,945 galaxies with $V_c > 100 \text{ km s}^{-1}$ and 1,021 galaxies with $V_c > 200 \text{ km s}^{-1}$. We estimate the corresponding numbers for the real Universe as follows. The observed density of galaxies is given by the field luminosity function (G. E., R. S. Ellis and B. A. Peterson, preprint). We assume that at all luminosities, 70% of galaxies are spirals with rotation velocities given by the Tully-Fisher relation^{26,27}, while 30% are ellipticals and SOs with circular velocities $3^{1/2}$ times the velocity dispersions obtained from the Faber-Jackson relation²⁸⁻²⁹. These assumptions predict 2,640 galaxies with $V_c > 100 \text{ km s}^{-1}$ and 615 galaxies with $V_c > 200 \text{ km s}^{-1}$, significantly fewer than found in our models. The discrepancy could be reduced in a variety of ways: the asymptotic circular velocity of haloes may overestimate the rotation of the galaxies which form within them; our *ad hoc* merging mechanism may miss a large number of mergers; our algorithm for identifying galaxies may be at fault, particularly at early times; or our assumed Hubble constant and initial fluctuation amplitude may need adjusting (this would require a re-evaluation of our previous work). We will not pursue this problem further here since it is peripheral to our main point. We intend to address it in a future paper using a more realistic treatment of galaxy formation and merging. It is curious that the abundance of dark haloes in our models is comparable to the observed abundance of galaxies; the 'maximal merging' assumption that each halo contains only one galaxy would thus give reasonable agreement with the observations (see also ref. 22).

Our major conclusions concern the spatial distributions of galaxies and mass in our models. Figure 1 presents spatial autocorrelation functions $\xi(r)$, for these distributions, with error bars derived from the model-to-model variation. The mass autocorrelation agrees well with our earlier work, verifying that the present simulations correctly match our models of larger regions. The galaxy autocorrelation functions are steeper than that of the mass and have a larger amplitude. Both effects are more pronounced for galaxies with larger circular velocities. Changing our merger criterion to $r_{\text{mer}} = 0$ or 200 kpc shifts the small-scale correlations up or down, but has very little effect for $\xi < 50$. A straight vertical shift of the mass curve in Fig. 1 shows that, over the separation range plotted, the mean correlation enhancement is about 1.8 for $V_c > 100 \text{ km s}^{-1}$ and 5 for $V_c > 250 \text{ km s}^{-1}$; for $V_c > 150 \text{ km s}^{-1}$ and $V_c > 200 \text{ km s}^{-1}$ (not plotted in the figure) the enhancements are 2.8 and 4 respectively. The distribution of galaxies is thus significantly biased with

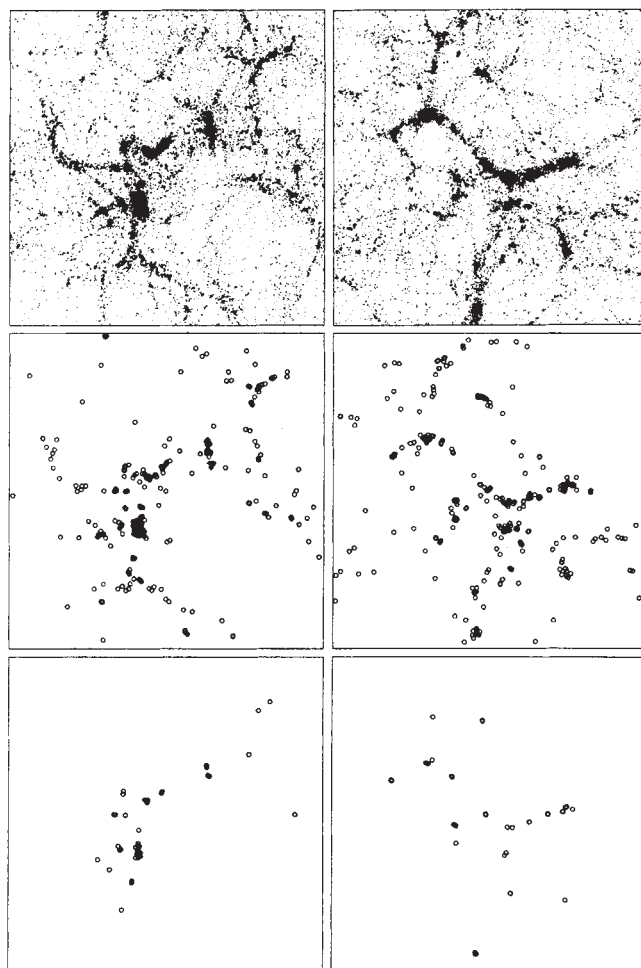


Fig. 2 Distribution of mass and 'galaxies' in two separate regions of one of our simulations. Each column shows projections of a slice of width $1/8$ of the simulated cube. The top row corresponds to the distribution of mass, while the middle and bottom rows display the distribution of 'galaxies' with $V_c > 100$ and 200 km s^{-1} , respectively.

respect to the mass, even for the smallest objects we can resolve. For $V_c > 250 \text{ km s}^{-1}$ the correlations are comparable to those of observed (bright) galaxies (represented in the figure by the dashed line of slope -1.8). As noted above, we expect to underestimate correlations on scales exceeding 5 Mpc because of the limited size of our simulations. We have also calculated the relative velocity dispersion of galaxy pairs in the simulations. The one dimensional dispersion is about 250 km s^{-1} at all separations, in good agreement with observations^{1,2}. It is interesting that the bias is still present if we assume only one galaxy per halo; central galaxies with $V_c > 200 \text{ km s}^{-1}$ have a correlation amplitude about 2.2 times that of the mass.

This bias arises because clumps in a high density region collapse earlier and accrete faster than similar objects in a low density region. As a result the typical mass and velocity dispersion of clumps is greater in protoclusters than in protovoids. For bright galaxies the natural bias we have discovered is strong enough to reconcile the dynamics of galaxy clustering with a flat universe. Even when averaged over the entire galaxy distribution, the bias will cause most standard estimates of the density of the universe to be low by factors of three or four.

It is instructive to make qualitative as well as quantitative comparisons of the distributions of mass and of galaxies. In Fig. 2 we give pictures of these distributions for two slices through our simulations. Each slice contains one eighth of the simulation volume. We plot separately the mass, galaxies with

$V_c > 100 \text{ km s}^{-1}$, and galaxies with $V_c > 200 \text{ km s}^{-1}$. The nature of the bias is quite clear in these pictures. The same structures are seen in each component, but the contrast of the structures is enhanced in the galaxy distributions. The enhancement increases with V_c . Thus voids in the bright galaxy distribution tend to be empty of fainter objects also, or at least to contain very few of them. The mass density in such regions is also low. In each of our three models we found a spherical region 900 to $1,200 \text{ km s}^{-1}$ in diameter which contain no galaxies with $V_c > 100 \text{ km s}^{-1}$. In each case, the average mass density in the sphere was less than 20% of the cosmic mean. Such voids are as large as could be expected in a simulation of size $2,500 \text{ km s}^{-1}$.

It has long been traditional to assume that galaxies will approximate a fair sample of the mass distribution in any universe where structure grows by hierarchical clustering. Our results show this assumption to be seriously in error for models of the kind required to obtain the observed properties of galaxy haloes and galaxy clusters. Haloes cluster more strongly than the mass, and this bias increases with the depth of the potential well of the haloes considered. Its effect is enhanced because more massive haloes often contain several galaxies. The bias depends on the assumptions that a galaxy condenses in every halo that forms, that the brightness of the galaxy reflects the depth of the halo potential, and that galaxy merging is not efficient enough to prevent cluster formation. There is no evidence for the predicted luminosity dependence of clustering in present data samples, although the range of luminosities in a magnitude limited galaxy sample is quite small. The observational situation is further complicated by the observed dependence of clustering on morphological type^{30,31}. The model presented here actually predicts the correlation functions of ellipticals and spirals to differ in the sense observed. It also suggests that voids in the distribution of bright galaxies should be underdense in lower mass objects. In a cold dark matter universe, natural bias may be strong enough to reconcile the observed kinematics and clustering of galaxies with the theoretical imperative of a flat universe.

This work was supported by NSF Presidential Young Investigator awards and by NATO, SERC and the Nuffield foundation. The computing was made possible by a generous allocation from the San Diego Supercomputing Center.

Received 12 August; accepted 25 September 1987.

- Davis, M. & Peebles, P. J. E. *Astrophys. J.* **267**, 465-482 (1983).
- Bean, A. J., Efstathiou, G., Ellis, R. S., Peterson, B. A. & Shanks, T. *Mon. Not. R. astr. Soc.* **205**, 605-624 (1983).
- Yahil, A., Walker, D. & Rowan-Robinson, M. *Astrophys. J.* **301**, L1-L5 (1986).
- Meiksin, A. & Davis, M. *Astr. J.* **91**, 191-198 (1986).
- Bardeen, J. M., Bond, J. R., Kaiser, N. & Szalay, A. S. *Astrophys. J.* **304**, 15-61 (1986).
- Davis, M., Efstathiou, G., Frenk, C. S. & White, S. D. M. *Astrophys. J.* **292**, 371-394 (1985).
- Rees, M. J. *Mon. Not. R. astr. Soc.* **213**, 75p-81p (1985).
- Rees, M. J. in *Clusters and Groups of Galaxies: Astrophys. Sp. Sci. Library*, vol. III (eds Mardirossian, F., Giuricin, G. & Mezzetti, M.) 485-497, (Reidel, Dordrecht, 1984).
- Dekel, A. & Silk, J. *Astrophys. J.* **303**, 39-55 (1986).
- Dekel, A. & Rees, M. J. *Nature*, **326**, 455-462 (1987).
- Blumenthal, G. R., Faber, S. M., Primack, J. R. & Rees, M. J. *Nature* **311**, 517-525 (1984).
- White, S. D. M., Frenk, C. S., Davis, M. & Efstathiou, G. *Astrophys. J.* **313**, 505-516 (1987).
- Peebles, P. J. E. *The Large Scale Structure of the Universe* (Princeton University Press, 1980).
- White, S. D. M. & Rees, M. J. *Mon. Not. R. astr. Soc.* **183**, 341-358 (1978).
- Silk, J. *Astrophys. J.* **211**, 638-648 (1977).
- Rees, M. J. & Ostriker, J. *Mon. Not. R. astr. Soc.* **179**, 541-552 (1977).
- Fall, S. M. & Efstathiou, G. *Mon. Not. R. astr. Soc.* **193**, 189-206 (1980).
- Toomre, A. in *Evolution of Galaxies and Stellar Populations* (eds Tinsley, B. M. & Larson, R. B.) 401-426 (Yale University Press, 1977).
- Fall, S. M. *Nature* **281**, 200-202 (1979).
- Peebles, P. J. E. *Astrophys. J.* **263**, L1-L5 (1982).
- Frenk, C. S., White, S. D. M., Efstathiou, G. & Davis, M. *Nature* **317**, 595-597 (1985).
- Frenk, C. S., White, S. D. M., Davis, M. & Efstathiou, G. *Astrophys. J.* (in the press).
- Efstathiou, G., Davis, M., Frenk, C. S. & White, S. D. M. *Astrophys. J. Suppl.* **57**, 241-260 (1985).
- Hockney, R. W. & Eastwood, J. W. *Computer Simulations Using Particles* (McGraw-Hill, New York, 1981).
- Huchra, J. P. in *Proc. European Southern Observatory Workshop on the Virgo Cluster* (eds Richter, O.-G. & Binggeli, B.) 181-200 (ESO Publications, Garching, 1984).
- Tully, R. B. & Fisher, J. R. *Astr. Astrophys.* **54**, 661-673 (1977).
- Aaronsen, M. & Mould, J. *Astrophys. J.* **265**, 1-17 (1983).
- Faber, S. M. & Jackson, R. E. *Astrophys. J.* **204**, 668-683 (1976).
- Davies, R. L., Efstathiou, G., Fall, S. M., Illingworth, G. & Schechter, P. L. *Astrophys. J.* **266**, 41-57 (1983).
- Davis, M. & Geller, M. J. *Astrophys. J.* **208**, 13-19 (1976).
- Giovannelli, R., Haynes, M. P. & Chincarini, G. L. *Astrophys. J.* **300**, 77-92 (1986).

Quasars of redshift $z = 4.43$ and $z = 4.07$ in the South Galactic Pole field

S. J. Warren*, P. C. Hewett*, P. S. Osmer† & M. J. Irwin*

* Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK

† Kitt Peak National Observatory, National Optical Astronomy Observatories, Tucson, Arizona 85725, USA

The detection of high-redshift quasars ($z > 3$) offers the most direct method for studying conditions in the Universe at early epochs. Analysis of the changes in the luminosity function of the quasar population places constraints on the epoch of formation of massive, gravitationally bound systems, while detailed observations of the quasar spectra provide information on the properties and evolution of intervening absorption systems such as the Lyman- α clouds. We report here the discovery of two new quasars of very high redshift; Q0051-279 of $z = 4.43$ and Q0101-304 of $z = 4.07$. These two quasars bring to five the total number of quasars known with $z > 4$ (refs 1-3). The redshift of Q0051-279 is the highest yet recorded and compares with the $z = 4.11$ of the previous most distant known quasar³. Both new quasars were found by the same multicolour technique and in the same UK-Schmidt-Telescope (UKST) field as Q0046-293, the first quasar found with a redshift $z > 4$ (ref. 1). The multicolour selection technique, coupled with other quantitative searches now in progress, provides the means for deriving a greatly improved description of quasar evolution at high redshifts.

The recent development of techniques that are capable of

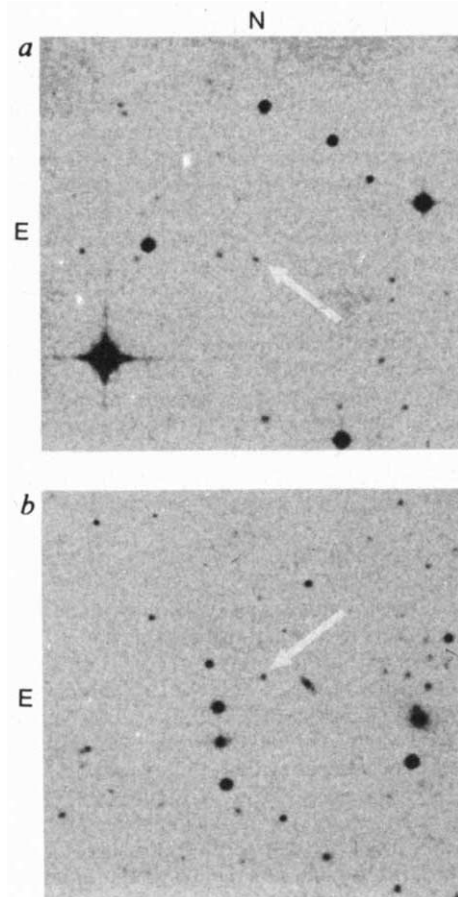


Fig. 1 Finding charts for a, Q0051-279 and b, Q0101-304, reproduced from direct plate R9594 (courtesy of UKST unit). The area is 5 arc min $\times$ 5 arc min. North is up and east is to the left.

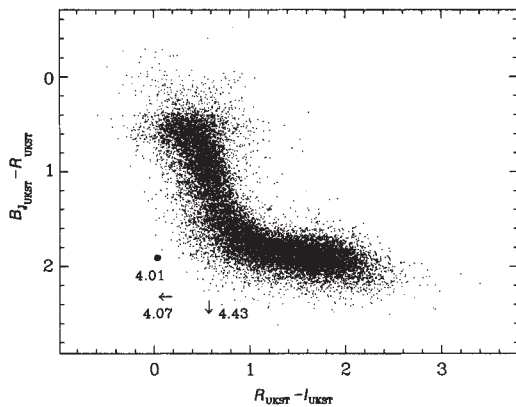


Fig. 2 $B_J - R$ against $R - I$ colour-colour diagram for stellar images brighter than $m_R = 20.0$ in the UKST field centred on 0 h 53 min, $-28^\circ 03'$ (SGP); a subset of the UB_VRI multicolour data set¹. The position of Q0046-293 ($z = 4.01$) is shown, together with the two new quasars of $z > 4$ described in this letter. The locations of the arrows correspond to the detection limits for the relevant pass-bands. The very large number of other objects missing from the I or B_J plates are omitted for clarity. The new quasars were selected in a density search of the full multicolour parameter space. The multicolour technique is described in detail in ref. 1. $\leftarrow$, Missing from both I plates; $\downarrow$, missing from both B_J plates.

finding high-redshift quasars in significant numbers, in surveys that cover large areas of sky and reach relatively faint limiting magnitudes, has led to a change in our understanding of how the quasar population evolves at early epochs. Based on earlier, primarily unsuccessful searches in small areas of sky, the existence of a significant decrease in the space density of quasars at high redshifts—the redshift cutoff—gained general acceptance. However, the detection of several quasars of redshift $z > 3.5$ by a number of groups has led to a reappraisal of the observational evidence. Indeed some recent analyses, based on the rather limited data available, claim the space density of bright quasars may actually be increasing at high redshifts^{4,5}, albeit at a much reduced rate compared to that at redshifts $z < 2$. This reappraisal is due, at least in part, to consideration of quasar evolutionary models that are not simple extrapolations of those which have successfully explained the observational data for redshifts $z < 2$.

In an investigation of a field of 30 square degrees centred on the South Galactic Pole (SGP) using a broad-band multicolour quasar selection technique¹, we have obtained spectra of 17 new quasars with $z > 3$, bringing to 24 the number of known quasars in this field with redshifts exceeding $z = 3.0$ and with magnitudes $m_R < 20.0$. Three of these 24 have redshifts exceeding 4.0. One of these three is Q0046-293 ($z = 4.01$) (see ref. 1). Here we describe the properties of the two new objects that have higher redshifts.

Q0051-279 was confirmed as a quasar of redshift $z = 4.43 \pm 0.01$ (position 0 h 51 min 49.8 s, $-27^\circ 58' 24''$ (equinox 1950.0)). The broad-band magnitudes (in the system defined by the emulsion and filter combinations used by the UKST, with B_J and V zero-pointed to the Johnson BV system, and R and I to the Kron-Cousins system) are $U > 21.2$ (absent from both U plates), $B_J > 22.4$, $V > 20.5$, $R = 20.0$, $I = 19.4$. The errors in the magnitudes are typically ± 0.1 for $m_B \approx 21$ and $m_R \approx 20$. Q0101-304 was confirmed as a quasar of redshift $z = 4.07 \pm 0.01$ (position 1 h 1 min 14.1 s, $-30^\circ 25' 4''$ (equinox 1950.0)). The broad-band magnitudes are $U > 21.2$, $B_J = 21.8$, $V = 20.1$, $R = 19.5$, $I > 19.4$. Finding charts for both objects are included in Fig. 1. Figure 2 shows the locations of the three quasars of $z > 4$ in a $B_J - R$ against $R - I$ colour-colour diagram. Both quasars are intrinsically bright, with absolute magnitudes $M_R \approx -28$ (for Hubble's constant $H_0 = 50 \text{ km s}^{-1} \text{ Mpc}^{-1}$ and deceleration parameter $q_0 = 0.5$).

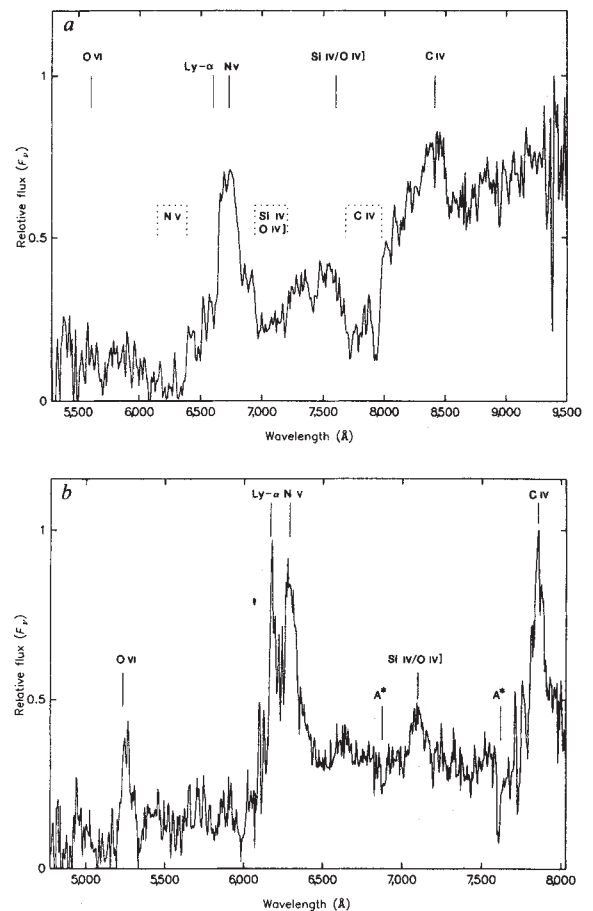


Fig. 3 Low-resolution spectra of *a*, Q0051-279 (25 Å) obtained in a 3,200 s exposure at the Anglo-Australian Telescope using the faint object red spectrograph, and *b*, Q0101-304 (11 Å) obtained in a 1,200 s exposure at the CTIO 4-m telescope using the Ritchey-Chretien spectrograph plus Air Schmidt camera. Relative flux per unit frequency is plotted versus wavelength in ångströms. The positions indicated for the major emission features are those corresponding to the adopted quasar redshifts. The three pairs of dashed lines for Q0051-279 show the approximate extent of the prominent broad-absorption-line systems associated with C IV, Si IV/O IV] and N V. The spectrum of Q0101-304 has not been corrected for atmospheric absorption, and features due to the atmospheric A and B bands are indicated by A*.

Low-resolution red spectra of Q0051-279 (obtained at the Anglo-Australian Observatory (AAO) on the night of 12/13 September 1987) and Q0101-304 (obtained at the Cerro Tololo Inter-American Observatory (CTIO) on the night of 19/20 August 1987) are shown in Fig. 3. Q0051-279 is clearly a member of the class of broad-absorption-line quasars and shows complex absorption troughs corresponding to C IV, Si IV/O IV] and N V. Absorption corresponding to Ly- α is probably also present, but the wavelength overlap of the N V absorption and any corresponding Ly- α system, coupled with the presence of the Ly- α forest precludes an unambiguous identification. The prominent red edges of the absorption features correspond to velocities of $\sim 15,000 \text{ km s}^{-1}$ blueward of the central emission peaks, with strong absorption present to velocities of $\sim 26,000 \text{ km s}^{-1}$ from the emission peaks. We have measured the redshift from the C IV ($\lambda = 1,549.1 \text{ Å}$) emission line alone. The Si IV/O IV] emission line is strongly affected by the absorption system corresponding to C IV, and the O VI line is completely absent. The Ly- α emission appears strongly absorbed, with the bulk of the observed feature at $6,700 \text{ Å}$ consistent with N V emission. Reliable estimates of line equivalent widths are not possible due to strong absorption. Similarly the slope of the

continuum to the red of Ly- α is uncertain, but it is apparently unusually steep. Q0101-304 exhibits features common in optically selected quasars: strong emission lines of Ly- α /N v (60 ± 7 Å), Si IV/O IV] (9 ± 3 Å) and C IV (46 ± 6 Å) are visible (the figures in brackets denote the rest-frame equivalent widths), and to the blue of the Ly- α /N v feature the continuum shows a significant depression due to absorption from the Ly- α forest. The O VI emission appears particularly strong, but a reliable measure of the equivalent width is impossible without higher-resolution data to define better the continuum in the Ly- α forest. We have measured the redshift from the C IV (1549.1 Å) and Si IV/O IV] (1,401.6 Å) emission lines.

The detection of three quasars of redshift $z > 4$ in a single UKST field demonstrates the potential for a quantitative study of the evolution of quasars at these high redshifts. Efforts to push the detection limits towards redshifts $z > 5$ over the next few years should allow a significant test of the predictions for galaxy formation of the popular cold dark matter models⁶.

We thank the staffs of the UKST, AAO (especially Steve Lee) and CTIO for their assistance, as well as Ed Kibblewhite and the members of the APM group. S.J.W. acknowledges the award of a SERC studentship. P.C.H. is supported by the Royal Society. The authors are visiting astronomers at Cerro Tololo Inter-American Observatory. CTIO and KPNO are operated by AURA, Inc. under contract with the US NSF.

Received 4 November; accepted 5 November 1987.

1. Warren, S. J. *et al.* *Nature* **325**, 131-133 (1987).
2. Schmidt, M., Schneider, D. P. & Gunn, J. E. *Astrophys. J.* **321**, L7-L10 (1987).
3. Hazard, C. *et al.* (in preparation).
4. Koo, D. C. & Kron, R. G. *Astrophys. J.* (in the press).
5. Hazard, C., McMahon, R. G. & Sargent, W. L. W. *Nature* **322**, 38-40 (1986).
6. Efsthathiou, G. & Rees, M. J. *Mon. Not. R. astr. Soc.* (in the press).

A non-thermal axially symmetric radio wake towards the galactic centre

Farhad Yusef-Zadeh* & John Bally†

* Laboratory for Astronomy and Solar Physics,
NASA/Goddard Space Flight Center, Greenbelt,
Maryland 20771, USA

† AT&T Bell Laboratories, Holmdel, New Jersey 07922, USA

We have discovered a highly unusual radio source lying within 1° of the galactic centre whose 'cometary' morphology suggests that it is a wake produced by a radio source moving supersonically with respect to the ambient interstellar medium. Most galactic sources of nonthermal radio continuum radiation are supernova remnants, which retain a degree of spherical symmetry. Observations of high sensitivity and resolution from the Very Large Array (VLA) radio telescope in New Mexico have shown that within a few degrees of the galactic centre the sky harbours a number of unusual asymmetric structures, including narrow filaments¹⁻³ and highly distorted supernova remnants⁴⁻⁷. Here we report the finding of a collimated axisymmetric nonthermal source.

The new source, G359.23-0.82, was discovered during an extensive radio continuum VLA survey of the galactic centre region at wavelengths of both 6 and 20 cm. We observe a linearly polarized compact ($\theta < 3$ arc s) core at the extreme eastern edge of a conical structure containing a highly collimated axial ridge which can be traced for over 12 arc min towards the west. The vertical extent of this source appears to narrow toward the western end. Observations between 2 and 20 cm show that the position angle of the axial ridge remains constant over a size scale from 10 arc s to 12 arc min. The conical envelope of the radio source shows several oscillations in its transverse extent, with length scale of 1.5 to 2 arc min. These oscillations resemble

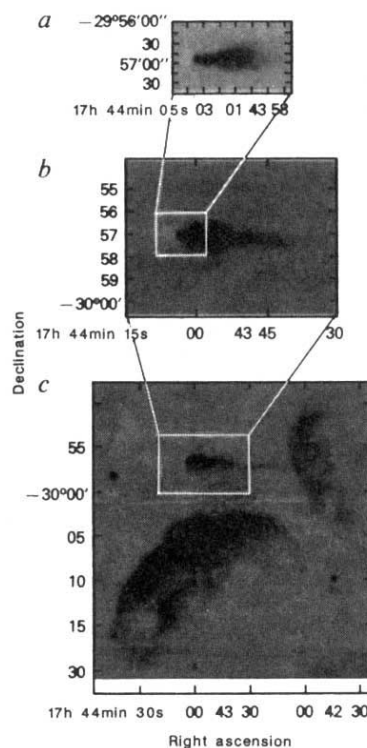


Fig. 1 The 20-, 6- and 2-cm radiographs of G359.23-0.82 and its surroundings. The bottom panel (c) is constructed by combining the 20-cm data corresponding to B/C and C/D arrays. This image is convolved by a gaussian beam whose full width at half maximum (FWHM) = $22 \text{ arc s} \times 22 \text{ arc s}$. The middle panel (b) presents the 6-cm image based on the C/D array. FWHM = $13.5 \text{ arc s} \times 11 \text{ arc s}$ position angle (PA = 55°). The top panel (a) shows the 2-cm image constructed from the D-array data. FWHM = $8.4 \text{ arc s} \times 3.2 \text{ arc s}$ (PA = -6°).

the axially symmetric periodic wakes seen in fluid flows behind a sphere or a disk⁸. This remarkable radio morphology, seen in three different radiographs at wavelengths of 2, 6 and 20 cm (Fig. 1a-c), is highly suggestive of a bow shock or a wake in the interstellar medium, produced either by the advance of the working surface of a hypersonic jet or by supersonic motion of a source with respect to the gas in the interstellar medium.

We observed G359.23-0.82 at 20-, 6- and 2-cm wavelengths using the VLA in its hybrid B/C, C/D and D-array configurations during October 1986, February 1987 and April 1987. In the hybrid configurations, the northern arm of the array is three times longer than the east and west arms, producing a nearly circular synthesized beam in a southerly direction towards the galactic centre. The 20- and 6-cm data were obtained with the B/C and C/D arrays respectively, so that the images at these wavelengths have the same synthesized beam size, permitting an accurate measurement of the spectral index. In each array we observed G359.23-0.82 five separate times for a total integration time of 40 min during each 8-h observing run, to insure relatively complete (u, v) coverage. We used 3C286 for flux calibration and 1748-253 and NRA0530 for antenna complex gains and instrumental polarization calibrations. Details of calibration and image processing will be given elsewhere.

Our 20-cm imaging of the field of view surrounding G359.23-0.82 (Fig. 1c) shows two shell-like non-thermal radio structures. One is located 10 arc min to the south, and is a portion of a previously unidentified supernova remnant (SNR) whose spectral and polarization characteristics will be published elsewhere. The other, located 12 arc min to the west, is the eastern portion

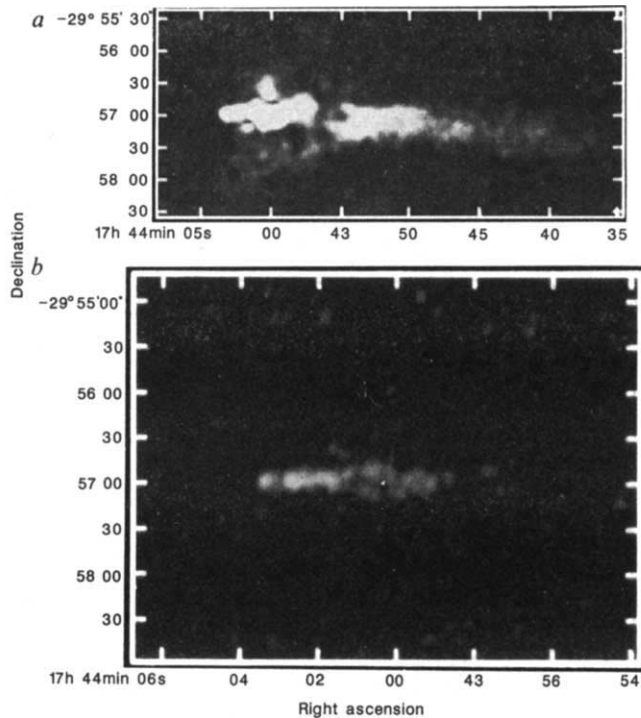


Fig. 2 Grey-scale presentations of the polarized intensity at 6 and 2 cm with resolutions similar to Fig. 1b, c.

of a remarkably circular shell-type SNR known as G539.2-0.5^{9,10}. The surface brightness-diameter relationship (Σ -D) relationship gives a physical diameter of 40 pc for this remnant, implying a distance of 8 kpc similar to the best estimates of the distance to the galactic centre. It is interesting to note that the tail end of our cometary radio source, G359.23-0.82, crosses the eastern edge of SNR G539.2-0.5, and that the axis of symmetry of the cometary object, when extended, intersects within 2 arc min of the centre of the SNR.

The cometary radio source G359.23-0.82 is highly polarized at both 6- and 2-cm wavelengths. No reliable polarized emission at 20 cm is detected from either G359.23-0.82 or the two supernova shells seen in the 20-cm field. Figure 2a and b show grey-scale representations of the 6 and 2-cm polarized intensity. These images show that the polarization peaks near the axis of symmetry of the radio source. Figure 3a and b shows the electric field distribution superimposed upon contour maps of the total and polarized intensity at 6 and 2 cm, respectively. The length of the line segments in Fig. 3a is proportional to the fractional linear polarization, which ranges from 5 to 60% of the flux at each point within the eastern portion of the source. The position angle of the electric field vectors changes abruptly by $\sim 90^\circ$ at several places along the major axis of the object. This behaviour is quasi-periodic, having the same length scale as the fluctuations in the transverse extent of the source. The presence of a high percentage of polarization demonstrates that most of the continuum emission is nonthermal synchrotron radiation and that the magnetic field is highly organized along the object. The magnetic field at minimum pressure and the synchrotron lifetime

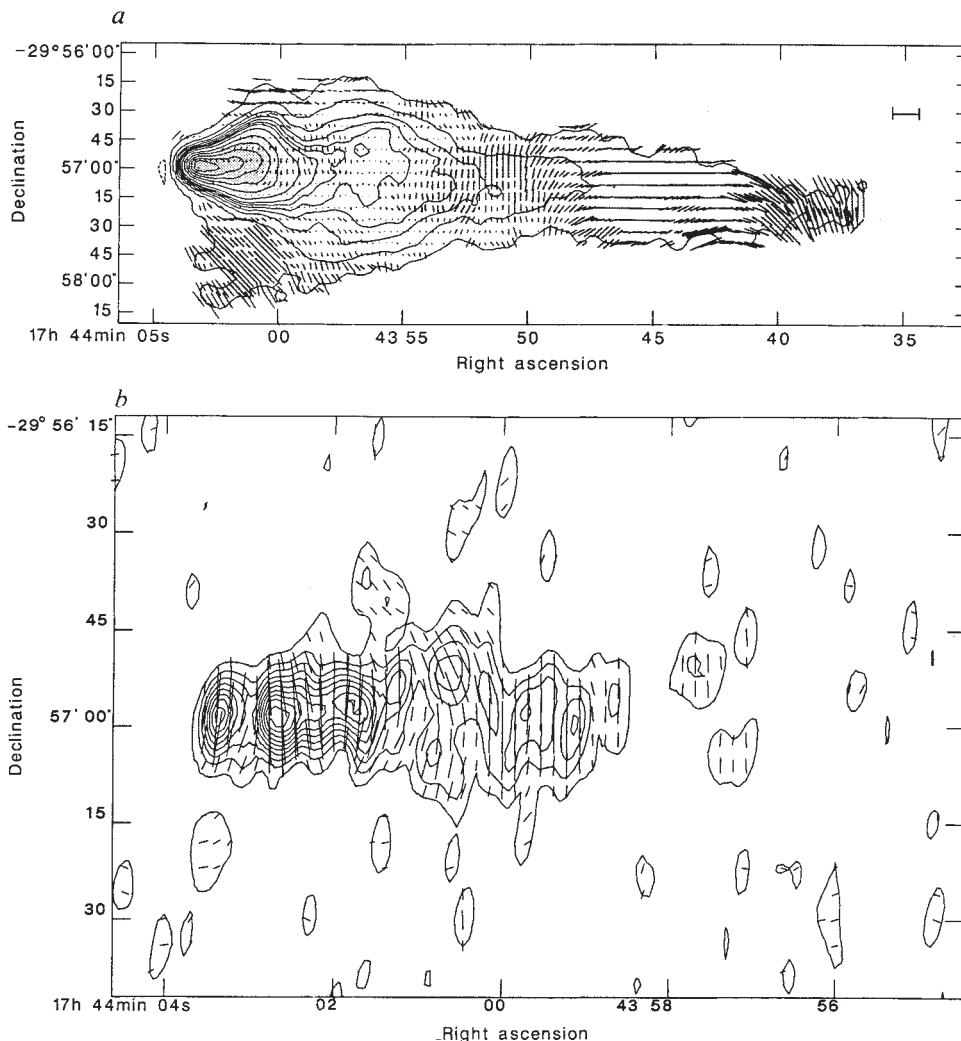
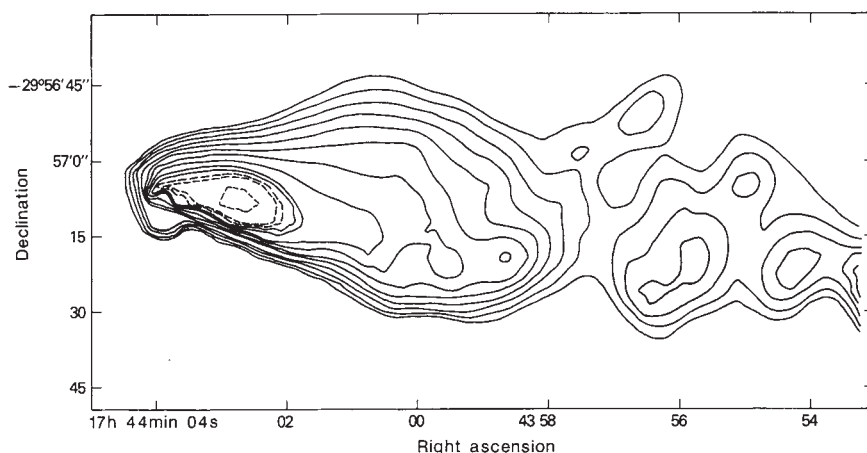


Fig. 3 Contour maps of the total and polarized intensity at 6 and 2 cm with resolutions FWHM = 9.6 arc s $\times$ 9.2 arc s (PA = 38 $^\circ$) and 8.4 arc s $\times$ 3.3 arc s (PA = -6 $^\circ$), respectively. Contour levels are set -2.5, -0.5, 0.5, 2.5, 5, 7.5, 10, 12.5, 15, 20, 30, 40, 50, 60, 70, 80 and 0.25, 0.5, 0.75, ..., 2.5 mJy per beam area. The superimposed line segments represent the directions of maximum electric field vectors with lengths proportional to fractional polarization at 6 cm and to polarized intensity at 2 cm. The bar in a corresponds to 100% of linear polarization. The fractional polarization measurements near the western edge of the source are contaminated by instrumental calibration errors introduced at positions several arc min from the field centre, which is near the eastern edge of the source.

Fig. 4 The spectral index map obtained from the 20- and 6-cm data, with $\text{FWHM} = 11 \text{ arc s} \times 10.3 \text{ arc s}$ ($\text{PA} = 56^\circ$). Contour levels correspond to spectral index $\alpha = 0.02, 0.01, 0.0, -0.1, -0.02, \dots, -0.3$. Solid lines represent negative contour levels.



for relativistic particles are estimated to be $\sim 4 \times 10^{-5} \text{ G}$ and $\sim 1.6 \times 10^6 \text{ yr}$.

Figure 4 shows a contour map of the spectral index between 6 and 20 cm. The spectral index (defined as α where $S_\nu \propto \nu^\alpha$) is flat, with $\alpha = 0.03$ at the position of the compact core located at the eastern edge of the radio source. The spectral index steepens away from the core, to a value $\alpha = -0.3$ at the western end of the source. This spectral index is consistent with a nonthermal origin of the radiation. Its variation suggests evolution of the population of radiating particles away from the optically thick compact source. Adiabatic expansion and synchrotron losses may account for the spectral index gradient.

Morphologically, G359.2–0.82 resembles a single lobe of very extended extragalactic radio sources¹¹, but neither the second lobe (expected if this were part of a double radio source) nor a candidate compact energy source (expected to be associated with an active galactic nucleus responsible for producing the radio double) has been seen. The head-tail morphology of this object resembles crudely the C-shaped appearance of narrow angle head-tail radio galaxies¹². However, this model requires a nearly 90° inclination angle. Furthermore, a flat spectrum and a high degree of polarization are uncharacteristic of head-tail sources. We argue that G359.23–0.82 is a galactic object, possibly located near the galactic centre, where several other morphologically peculiar radio sources have been recently discovered^{4–6}, notably the axisymmetric nonthermal galactic radio source G357.7–0.1, which shows similar morphology to the cometary source described here.

The total flux of G359.28–0.82 at 1.4 GHz is $\approx 2 \text{ Jy}$, and the radio total luminosity is $L_R \approx 2.6 L_\odot$, assuming that it is at a distance of 8.5 kpc, near the galactic centre. This is comparable to the radio luminosity of shell SNRs, and about an order of magnitude less than that of Crab Nebula-like 'plerionic' remnants¹³.

The apparent opening angle of the wake measured from the apex containing the compact core is $\sim 10^\circ$ (Fig. 1c). Assuming that the conical shape is a bow shock produced by supersonic motion through the interstellar medium of a source at the apex, we can crudely estimate the needed velocity. The opening angle θ of a bow shock is related to the velocity of the source V and the sound speed C by $V = 2f(\phi)C/\theta$, where ϕ is the angle between V and the line of sight, and $f(\phi)$ is a correction to account for projection effects. Apparently the apex is advancing at $\sim 5C$ through the ambient interstellar medium. The overall length of the wake is about 12 arc min or 30 pc at the distance of the galactic centre. Assuming that the energy source of the apex has a sufficiently long lifetime τ that it can traverse this distance, we find $\tau(\text{yr}) = X/V = X\theta/(2f(\phi)C) \approx 2.6 \times 10^4 (\theta/10^\circ)(C/100 \text{ km s}^{-1})^{-1}$. There are two stable phases in the interstellar medium, a hot phase with temperature $T > 10^6 \text{ K}$ ($C \approx 100 \text{ km s}^{-1}$) and a cold phase with $T < 10^4 \text{ K}$ ($C \approx 10 \text{ km s}^{-1}$). In the cold phase $\tau \approx 2.6 \times 10^5 \text{ yr}$. For the

stable hot phase, the transverse velocity is about 300 km s^{-1} , and for the cold phase $V < 30 \text{ km s}^{-1}$. The most reasonable model is that G359.2–0.82 is evolving in a region occupied by the hot phase of the interstellar medium. This is consistent with the absence of molecular and ionized H II gas which can be associated with this source.

We discuss two possible models which might explain the wake: in the first, the apex is the working surface of a highly supersonic low-density jet, whose source is located more than 12 arc min due west. In this view, this source may harbour an object similar to ScoX1¹⁴ or SS433¹⁵ which produce oppositely directed, highly collimated particle beams. The energy source may be an accreting compact object in a binary system, and may lie at the centre of the shell-like remnant G359.1–0.5. The object would then be similar to W50, the extended non-thermal source surrounding SS433. Unfortunately, there is no evidence for a counter jet. It is possible that radio emission is produced near the working surface of one jet only or that the jet is truly one-sided. However, there is no candidate for the source from which the jet originates.

In our second model, the compact source near the apex of G359.23–0.82 contains the dominant source of energy, which is moving at high velocities relative to the ambient medium. Such a source might be either a rapidly spinning neutron star of the type thought to be inside 'Crab-like' plerions, or a binary system containing at least one compact object. The high relative velocity between the ambient medium and the energy source may arise because the source is a member of a class of high velocity objects (pulsars are known to have velocities up to 250 km s^{-1} ; high velocity stars or extreme population II sources may also be common near the galactic centre) or because the interstellar gas is moving supersonically with respect to the source (the gas near the galactic centre is known to have peculiar motions of up to 150 km s^{-1}).

It has been suggested that a runaway binary, in which a compact object is formed by a supernova explosion, may power some highly asymmetric radio sources^{6,16–18}. The entire binary system may gain a large velocity as a consequence of the supernova event. Only plerionic supernova remnants powered by a rotating neutron star, or compact objects in accreting binary systems, producing X-ray emission or powerful, often relativistic particle beams, have luminosities capable of powering this wake. HI absorption study plus high-resolution radio-continuum and X-ray observations of this object are essential to further our understanding of the nature of G359.23–0.82.

We thank C. O'Dea for useful discussions. F.Y.-Z. is a Resident Research Associate of the National Research Council. VLA is operated by Associated Universities Inc. for the National Science Foundation.

Received 10 July; accepted 2 September 1987.

1. Yusef-Zadeh, F., Morris, M. & Chance, D. *Nature* **310**, 557–561 (1984).
2. Morris, M. & Yusef-Zadeh, F. *Astrophys. J.* **90**, 2511–2513 (1985).

3. Liszt, H. *Astrophys. J.* **293**, L65–L67 (1985).
4. Shaver, P. A., Salter, C. J., Patnaik, A. R., van Gorkom, J. H. & Hunt, G. C. *Nature* **313**, 113–115 (1985).
5. Becker, R. H. & Helfand, D. J. *Nature* **313**, 115–118 (1985).
6. Helfand, D. J. & Becker, R. H. *Nature* **313**, 118–119 (1985).
7. Caswell, J. L. *et al. Mon. Not. R. astr. Soc.* **225**, 329–334 (1985).
8. Birkhoff, G. & Zaranonello, E. H. *Jets, Wakes and Cavities* Ch. 8 (Academic, New York, 1957).
9. Downes, D., Goss, W. M., Schwartz, U. & Wouterloot, J. G. A. *Astr. Astrophys. Suppl.* **35**, 1–22 (1984).
10. Reich, W. & Furst, E. *Astr. Astrophys. Suppl.* **57**, 165–167 (1984).
11. Wilson, A. S. & Ulvstad, J. S. *Astrophys. J.* **275**, 8–14 (1983).
12. O'Dea, C. & Owen, F. *Astrophys. J.* **301**, 841–859 (1986).
13. Weiler, K. W. in *The Crab Nebula and Related Supernova Remnants* (eds Kafatos, M. C. & Henry, R. B. C.) 265–285 (Cambridge University Press, 1985).
14. Fomalont, E. B., Geldzahler, B. J., Hjelm, R. M. & Wade, C. M. *Astrophys. J.* **275**, 802–807 (1983).
15. Margon, B., Grandi, S. & Downes, R. A. *Astrophys. J.* **241**, 306–315 (1980).
16. Gott, J. R. *Astrophys. J.* **173**, 227–234 (1972).
17. Clark, D. H., Parkinson, J. H. & Caswell, J. L. *Nature* **254**, 674–676 (1975).
18. Haynes, R. F. *et al. Nature* **324**, 233–253 (1986).

Interstellar shock waves and ^{10}Be from ice cores

C. P. Sonett*, G. E. Morfill† & J. R. Jokipii‡*

* Department of Planetary Sciences and Lunar and Planetary Laboratory, University of Arizona, Tucson, Arizona 85721, USA

† Max-Planck Institute for Extraterrestrial Physics, Garching bei München, FRG

‡ Department of Astronomy and Steward Observatory, University of Arizona, Tucson, Arizona 85721, USA

The anomalously high concentrations of ^{10}Be in Antarctic ice cores^{1,2}, uncorrelated with $\delta^{18}\text{O}$, are consistent with an increase in the atmospheric cosmic ray (CR) flux from CR acceleration in propagating interstellar shock waves, which envelop the heliosphere and whose source may be ancient supernovas. This mechanism is an alternative to the model where decreases in geomagnetic field intensity, associated with geomagnetic reversals, periodically enhance the CR flux at the top of the atmosphere. That CR variations attributable to interstellar events such as supernova shock waves have so far not been observed in the CR record is a long-standing issue^{3–5}. If our interpretation of the ^{10}Be spikes is correct, it marks the first such observation. The possibility that 'direct' observation of the interstellar medium can be made using the ^{10}Be record would be an important adjunct to the study of cosmic rays in the interstellar medium.

Steep and relatively brief increases in the concentration of ^{10}Be are noted in ice cores from both the Vostok and the Dome C Antarctic drilling sites. Two major variations (increases) are found at ~33 and 60 kyr¹ BP (before present), using the time/ice-thickness model of Lorius *et al.*⁶. The complete 150-kyr record of ^{10}Be in the Vostok ice core for the most part corresponds closely to the $\delta^{18}\text{O}$ record, but the two ^{10}Be 'spikes' appear to be independent of the corresponding $\delta^{18}\text{O}$ record, which is thought to be controlled by temperature above the tropospheric inversion layer⁷. Figure 1, adapted from Raisbeck *et al.*¹, shows these uncorrelated spikes (marked by D_1 and D_2) as part of the total ^{10}Be and $\delta^{18}\text{O}$ records. From Fig. 1 of Raisbeck *et al.*¹ it can be seen that from 20 to 160 kyr BP the time axis is linear, although from 0 to 20 kyr BP the relation between time and ice-depth may be more complicated. In Fig. 1 of this paper a linearized timescale is determined by assuming local linearity from 0 to 20 kyr BP different from that for the remainder of the graph. The two different linearities are unimportant in the rest of this paper, because the data before 20 kyr BP are used only to establish an initial value from the earlier timescale. Figure 2a includes an expanded view of both D_1 and D_2 superimposed on a common relative timescale. The appearance of the two spikes suggests two distinct events; our discussion centres on D_1 .

In our interstellar model for the spikes, we follow the argument of Raisbeck *et al.*¹ for a CR flux increase as the cause of

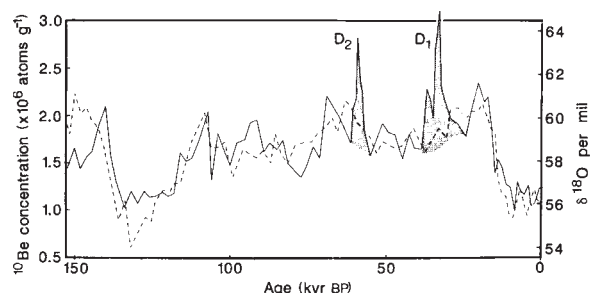


Fig. 1 Concentrations of ^{10}Be (solid line) and $\delta^{18}\text{O}$ (broken line) from 150 kyr BP to the present. Spikes marked by D's are hypothesized interstellar CR shock signatures. The time base in this figure is a linearized version of the time axis used by Raisbeck *et al.*¹ (see text). (Figure adapted from Raisbeck *et al.*¹.)

the beryllium enhancements. Then the increased concentrations of ^{10}Be in the Vostok ice core, reflecting the atmospheric inventory, suggest three main possibilities: (1) unexpectedly low values of the interplanetary magnetic field over a long period of time, due to low solar activity; (2) anomalous decreases in the magnitude of the geomagnetic field, leading to a greater CR flux at the top of the atmosphere; or (3) an interstellar source of enhanced CR. A doubling of the CR flux at energies near 1 GeV is a possible consequence of changes in the solar modulation of the GCR (galactic) component (compare Fig. 8 of ref. 4). As noted below, the characteristic lifetime of the spikes is several kyr. Most solar dynamical processes have much shorter timescales, and there is difficulty with invoking solar variability to produce the characteristic spikiness. Arbitrarily ruling out a depressed solar period on the basis of existing data would be a tautological exercise, but what evidence exists is negative. The discussion which follows emphasizes that a key issue in assigning a model for the beryllium spike is the relative merit of models based on geomagnetic reversals or excursions (model 2) or interstellar sources (model 3).

Both ^{14}C and ^{10}Be are produced by nuclear processes (reaction and spallation respectively) in interactions of GCR with major atmospheric constituents, and both are affected by changes in the intensity of the Earth's magnetic field^{8–10}. Global mixing of the radiocarbon inventory is complete over times variously estimated to lie between 10 and 100 yr¹¹. For ^{10}Be the relative timescales for precipitation and mixing suggest that the 30% of ^{10}Be estimated to be produced in the troposphere is precipitated out rapidly; that part of the record would thus be of local origin, tending not to reflect any global effect such as a geomagnetic reversal. Furthermore, because the Vostok site is near the south magnetic dip pole, and because ^{10}Be is produced by spallation of high GeV (GCR) on the major atmospheric constituents, the GCR flux generated in the troposphere above the Vostok site should be relatively unaffected by a geomagnetic cutoff. On the other hand, the remaining 70% of ^{10}Be , generated in the stratosphere, is thought to have a residence time of about 1 yr. For that component, composed partly of Be produced by mid- and low-latitude stratospheric CR subject to a variable geomagnetic cutoff, the polar precipitation might be affected by a globally changing geomagnetic field. Raisbeck *et al.*¹² have shown a correspondence between the Brunhes–Matuyama reversal 730 kyr ago and a doubling from about 3×10^9 ^{10}Be atoms g^{-1} to a peak of about 6×10^9 atoms g^{-1} . Lal¹³ gives a current discussion of atmospheric and polar ^{10}Be .

In considering the possibility that the anomalies are due to a sudden change in the geomagnetic field, we should distinguish between true field reversals and the excursions alluded to above, which are unexpected decreases in the field with a return to an earlier larger value but without a change in polarity. These are arguably local, identified with anomalies restricted to perhaps 200–500 km in extent¹⁴. Of the seven reversals within the

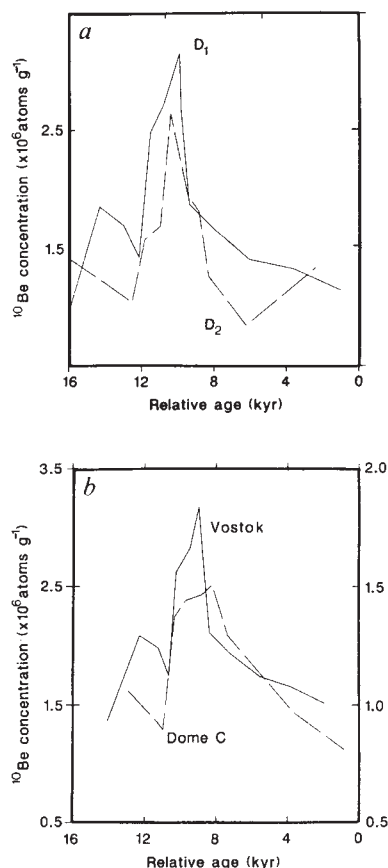


Fig. 2 Expanded view (a) of D_1 (solid line) and D_2 (broken line). The time axis is relative to the peaks are separated by some 26 kyr. The vertical axis is absolute ^{10}Be concentration. The time shift is to facilitate comparison of the spike shapes. (From data of Raisbeck *et al.*¹.) Expanded view (b) of D_1 , observed at the Vostok site (solid line), and the same event detected at Dome C (broken line). There is some evidence that the larger spike contains significant features which might be attributed to a more complex shock structure than just a simple wave, but on the other hand may reflect terrestrial (atmospheric) modification of the ^{10}Be precipitation. Timescales are shifted slightly for convenience of superposition. Left ^{10}Be scale corresponds to D_1 (solid line) and right scale (broken line) to D_2 . (From data of Raisbeck *et al.*¹.)

past 50 kyr tabulated by Smith¹⁴ (see also Creer¹⁵) the Laschamp/Olby event is the most conspicuous, but even so its age is uncertain, with estimates ranging from 8 to 50 kyr BP. The weightiest evidence against the reality of this one prominent reversal comes from the Lac du Bouchet sediments, which disclose no reversal. We cannot unequivocally rule out a reversal for the time period of interest for the ^{10}Be spikes, but as the Lac du Bouchet sediments show no reversal even under their ideal conditions, there is significant doubt as to the reality of a reversal near these times (K. M. Creer, personal communication). Additional evidence against reversals as a source of the Be spikes is the appearance of two spikes, and their closeness in time to each other and to the well-documented Brunhes-Matuyama reversal. Palaeomagnetically defined reversals of the Earth's magnetic field suggest a mean episodic interval of $\sim 10^7$ yr^{16,17}. It is statistically unlikely that three such events would have taken place within less than the 7×10^6 yr between Brunhes-Matuyama and now.

Figure 2b shows expanded profiles of D_1 from both the Vostok and Dome C stations. Rise and fall slopes for the steepest parts of the records are about 10^4 ^{10}Be atoms $\text{g}^{-1} \text{m}^{-1}$ for the Vostok record and about 5×10^3 ^{10}Be atoms $\text{g}^{-1} \text{m}^{-1}$ for Dome C. The former has a thickness between the steepest slopes of 27 m

($\sim 1,900$ yr) and the latter 21 m ($\sim 1,500$ yr). In view of the proximity of the two stations and the short atmospheric 'wash-out' time, the significant differences in spike thickness are conjectured to arise from variations in ice dynamics. This makes an exact measure of spike thickness impossible, but for the purposes of this report a qualitative account suffices. Both D_1 and D_2 (Fig. 2a) display a slight asymmetry in rise and fall times, with the fall being slightly slower, typical of shock waves. Both spikes show a characteristic 'tail', conjectured to correspond to the tail-off in CR flux behind the shock wave. Interpretation in terms of flux tubes containing an enhanced CR intensity is also possible, but we think it unlikely that static structures containing enhanced CR can survive long enough to be observed.

The envelopment of the heliosphere by an interstellar shock wave decreases its radius. For a heliospheric extent of 100 AU and a shock speed of 300 km s^{-1} , the decrease cannot occur in less than ~ 1.5 yr or more, depending upon the gradient of momentum in the leading edge of the shock wave and the dynamics of the solar wind. The CR flux increase within the shock wave is distributed through the heliosphere in a time determined by diffusion and drift, both of which are very short compared to the characteristic times of D_1 and D_2 . Assuming that D_1 is an enhancement of ^{10}Be due to an interstellar shock wave, the age of the putative parent supernova explosion is found from the width of the enhancement and the adiabatic cooling time (using an approximate Sedov solution¹⁸) $\tau_{\text{ad}} = r_s/3v_s$, where $v = 3v_s r/4r_s$ is the gas expansion velocity behind the shock. Because energetic particles diffuse slowly behind the shock, their evolution is governed by convection and cooling. The speed of particle convection away from the shock is $\Delta v = v_s(1 - 3r/4r_s)$, and the associated scale length for the decrease in CR intensity inside the post-shock remnant is approximated by $r_s/12$. Detailed numerical modelling yields similar values¹⁹, as do approximate analytic solutions. As $\Lambda = v_s t_s$ (where t_s is the shock (spike) width at $1/e$ maximum), $r_s/v_s = 12t_s$, and from the Sedov blast wave solution $r_s = at^{2/5}$ (where t is the age of the blast wave and $a = (6.25E/\pi\rho)^{1/5}$ is a constant depending on the initial blast wave energy E and the density of the interstellar medium ρ). Then $v_s = 0.4at^{-3/5}$, $r_s/v_s = 5t/2$ and the shock age $t = 4.8t_s$. We evaluate t_s using the solid line of Fig. 2b. If we regard the initial smaller peak as unrelated to the main peak, then $t_s = 2-4$ kyr. Alternatively, if the smaller peak contributes to the main peak, with the intervening dip taken to be an artefact, $t_s = 8$ kyr. The onset of D_1 was ~ 33 kyr ago so the age of the explosion is $33 + 4.8t_s$ kyr or ~ 45 kyr ago, with a total uncertainty of ~ 10 kyr. The observed lack of enhancement of cosmogenic nuclides from the meteorite record is consistent with the narrowness of the spikes. From the factor of two ^{10}Be enhancement at D_1 , it is possible to calculate the acceleration efficiency, ϵ . This is estimated as $0.01 < \epsilon < 0.13$ for $10^{50} \text{ erg} < E < 10^{51} \text{ erg}$ and a local interstellar density $10^{-26} \text{ g cm}^{-3} < \rho < 10^{-25} \text{ g cm}^{-3}$. These acceleration efficiencies agree with experimental values obtained at the Earth's bow shock²⁰.

It is interesting to compare the results obtained here with the implications drawn about the local interstellar medium on astronomical evidence. Cox and Anderson²¹ suggest that the Sun is inside a supernova remnant which is $\sim 10^5$ yr old, and was produced by a supernova of energy $E = 5 \times 10^{50} \text{ erg}$ expanding into a medium of density $\rho = 0.8 \times 10^{-26} \text{ g cm}^{-3}$. (For implications regarding local ^{26}Al production, see Morfill and Hartquist²².) As previously noted, the spike D_1 , if interpreted as a result of a supernova shock, would place the event at about 4.5×10^4 yr ago. The small difference in age estimated by Cox and Anderson and by us can be accounted for by small adjustments in ρ and E and perhaps also by uncertainty in the shock thickness estimate.

Finally we comment on the probability of detecting a supernova shock wave. The probability of a shock of speed v_s in the range $v_1 < v_s < v_2$ enveloping the solar system in time T is given

by $P = \dot{N}TV/V_g$ where V_g is the galactic volume, $\dot{N}$ the supernova rate, and V the galactic volume available to a shock wave within the specified velocity range upon solar system encounter. Then $P = 4\dot{N}TV[1 - v_1^2/v_2^2]/3\rho V_g v_1^2$. For $v_1 = 100 \text{ km s}^{-1}$, $v_2 = 600 \text{ km s}^{-1}$, $\dot{N} = 0.03 \text{ yr}^{-1}$, and $V_g = 7 \times 10^{10} \text{ pc}^3$, $P = 0.3 T_{1.5} E_{50}/\rho_{-26}$ where $T_{1.5}$ is time in units of 150 kyr, E_{50} is the supernova energy in units of 10^{50} erg , and ρ_{-26} is density in units of $10^{-26} \text{ g cm}^{-3}$. Thus there is a 30% probability of 'catching' such an event during the observation period. This makes more credible the possibility that both D_1 and D_2 are interstellar shock waves.

We thank A. J. T. Jull, J. Beer and U. Siegenthaler for discussion of the ^{10}Be problem, K. M. Creer for information on possible recent geomagnetic reversals, and G. M. Raisbeck for clarification of the precipitation process and manuscript errors. C.P.S. and J.R.J. are supported by the Solar-Terrestrial program of NSF, and J.R.J. is also supported by the NASA High-energy Astrophysics program. This work was done while G.M. was visiting the University of Arizona and partly while C.P.S. was visiting the University of Edinburgh.

Received 18 May; accepted 20 September 1987.

1. Raisbeck, G. M. *et al. Nature* **326**, 273-277 (1987).
2. Yiou, F. *et al. Nature* **316**, 616-617 (1985).
3. Axford, W. I. *Proc. 17th Int. Cosmic Ray Conf.* **12**, 155-175 (1987).
4. Jokipii, J. R. & Marti, K. in *The Galaxy and the Solar System* (eds Smolukoski, R. & Bahcall, J. N.) 116-128 (Univ. Arizona, Tucson, 1986).
5. Bhat, C. L., Houston, B. P., Meyer, C. J. & Wolfendale, A. W. *Proc. 19th Int. Cosmic Ray Conf.* **1**, 345-347 (1985).
6. Lorus, C. *et al. Nature* **316**, 591-596 (1985).
7. Raisbeck, G. M. *et al. Geophys. Res. Lett.* **8**, 1015-1018 (1981).
8. Buchta, V. *Nature* **224**, 681-683 (1969).
9. Barton, C. E., Merrill, R. T. & Barbetti, M. *Phys. Earth planet. Int.* **20**, 96-115 (1979).
10. Sonett, C. P. *Rev. Geophys. Space Phys.* **22**, 239-254 (1984).
11. Houtermans, J. C., Suess, H. E. & Oeschger, H. *J. Geophys. Res.* **78**, 1897-1908 (1973).
12. Raisbeck, G. M., Yiou, F., Bourles, D. & Kent, D. V. *Nature* **315**, 315-317 (1981).
13. Lal, D. *Geophys. Res. Lett.* **14**, 785-788 (1987).
14. Smith, G. *thesis*, Univ. Edinburgh (1985).
15. Creer, K. M. *Geophys. Surveys* **7**, 125-160 (1985).
16. Raup, D. M. *Nature* **314**, 341-343 (1985).
17. Lowrie, W. & Ogg, J. C. *Earth planet. Sci. Lett.* **76**, 341-349 (1986).
18. Sedov, L. I. *Similarity and dimensional methods in mechanics* (GITTL, Moscow, 1957); English translation (Academic, New York 1959).
19. Jokipii, J. R. *Astrophys. J.* (submitted).
20. Möbius, E., Scholer, M., Schopke, N., Lüder, H., Paschmann, G. & Hovestadt, D. *Geophys. Res. Lett.* (submitted).
21. Cox, D. P. & Anderson, P. R. *Astrophys. J.* **253**, 268-289 (1982).
22. Morfill, G. E. & Hartquist, T. W. *Astrophys. J.* **297**, 194-198 (1985).

Possible relevance of soliton solutions to superconductivity

T. D. Lee

Department of Physics, Columbia University, 538 West 120th Street, New York, New York 10027, USA

The newly discovered high-temperature superconductors¹ indicate that there might exist another underlying mechanism for superconductivity, different from the BCS theory². Recent suggestions of bipolarons^{3,4} and resonant valence bonds⁵, among others (see, for example, refs 6,7), point to the importance of an effective complex boson field $\phi(\mathbf{x})$. For example, $\phi(\mathbf{x})$ could be formed from a pair of electron fields $\psi_\uparrow$ and $\psi_\downarrow$ of opposite spin: $\phi(\mathbf{x}) \equiv \langle \psi_\uparrow(\mathbf{x} + \mathbf{l}) \psi_\downarrow(\mathbf{x} - \mathbf{l}) \rangle_l$ where $\langle \rangle_l$ denotes the appropriate average over the spacing l . The corresponding phenomenological Lagrangian density may be written as

$$\mathcal{L} = \frac{\partial \phi^\dagger}{\partial t} \frac{\partial \phi}{\partial t} - \nabla \phi^\dagger \cdot \nabla \phi - u(\phi^\dagger \phi) \quad (1)$$

where as $|\phi| \rightarrow 0$,

$$u \rightarrow m^2 \phi^\dagger \phi + g(\phi^\dagger \phi)^2 + \dots \quad (2)$$

with m acting as the effective mass and g as a nonlinear coupling. The replacement of ∂_μ (where μ denotes the space-time indices) by $\partial_\mu - 2eA_\mu$ gives the modification when there is an electromagnetic field A_μ . In models that assume a relatively tight binding between the electron pair, one might expect the coefficient m^2 to be positive and insensitive to the temperature T (for T less than the transition temperature T_c), in contrast to that in the standard Ginzburg-Landau equation⁸ (which has m^2 negative and proportional to $T - T_c$). This difference changes the character of the solution, making the customary acquisition of a non-zero $|\phi|$ through the Higgs mechanism inapplicable. The purpose of this note is to call attention to the property that, when $m^2 > 0$ and if there exists some attractive nonlinear interaction, the low-energy states of equation (1) may be dominated by solitons, instead of by the usual plane-wave solutions of particles of mass m .

Define $\varepsilon \equiv u - m^2 \phi^\dagger \phi$, which contains all the nonlinear interactions of ϕ . We assume that, at least for a limited range of $|\phi|$, ε is negative; that is,

$$\varepsilon < 0 \text{ for } \sigma_1 < |\phi| < \sigma_1 + \delta \quad (3)$$

where σ_1 and δ are both positive. (If $g < 0$ in equation (2), then

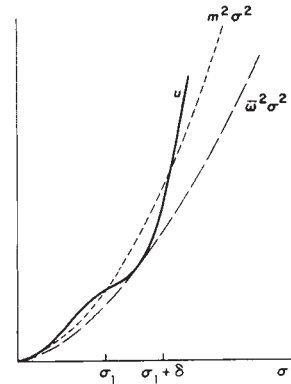


Fig. 1 The solid line gives an example of $u(\sigma) = m^2\sigma^2 + \varepsilon(\sigma)$, where $\varepsilon(\sigma) < 0$ for $\sigma_1 < \sigma < \sigma_1 + \delta$; the short-dashed line is $m^2\sigma^2$ and the long-dashed line is $\omega^2\sigma^2$, which is tangent to $u(\sigma)$.

$\sigma_1 = 0$.) As we shall see, when the space-dimension D is 1, independently of the behaviour of ε outside the range given in equation (3) and no matter how small ε and δ are within that range, there always exist non-topological soliton solutions. Such solutions can also exist in any higher dimension $D > 1$, but there is a condition on ε and δ ; this condition becomes more stringent as D increases. Furthermore, these solutions are valid in classical physics as well as in quantum mechanics⁹.

Consider first $D = 1$ and assume $\phi = \sigma(x)e^{-i\omega t}$. From equation (1) it follows that $\sigma(x)$ satisfies $\omega^2\sigma + d^2\sigma/dx^2 - \frac{1}{2}du/d\sigma = 0$, which, together with the boundary condition $\sigma = 0$ at $|x| = \infty$, gives

$$\left(\frac{d\sigma}{dx}\right)^2 - u + \omega^2\sigma^2 = 0 \quad (4)$$

This is analogous to the equation of energy conservation for a non-relativistic particle of 'position' σ and 'time' x moving in a 'potential' $v(\sigma) \equiv \omega^2\sigma^2 - u(\sigma)$. Both $u(\sigma)$ and $v(\sigma)$ are illustrated in Figs 1 and 2. Because of expression (3), there exists an $\omega^2 < m^2$ such that the parabola $\omega^2\sigma^2$ is tangent to $u(\sigma)$, as shown in Fig. 1. Hence, for any ω^2 between ω^2 and m^2 , there exists a soliton solution given by

$$x - x_0 = \pm \int_{\sigma_0}^{\sigma} v(\sigma)^{-1/2} d\sigma \quad (5)$$

as shown in Fig. 3. When $x \rightarrow \pm\infty$, $\sigma \approx \exp[-(m^2 - \omega^2)^{1/2}|x|]$; this gives a correlation length $(m^2 - \omega^2)^{-1/2}$ which approaches

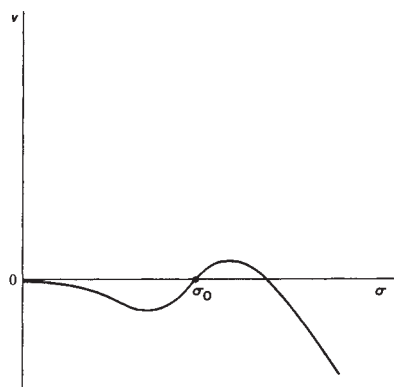


Fig. 2 Schematic plot of $v(\sigma) = \omega^2\sigma^2 - u(\sigma)$, for $\bar{\omega} < \omega < m$.

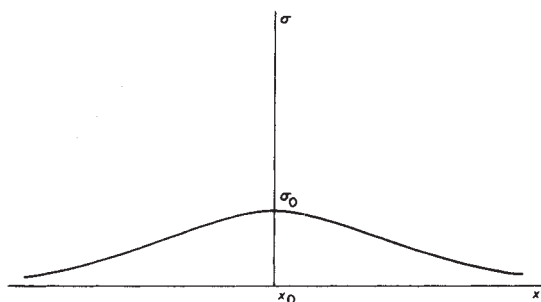


Fig. 3 Example of a non-topological soliton, thus called because as $|x| \rightarrow \infty$, $|\phi| \rightarrow 0$ (in any dimension), where $\phi = \sigma e^{-i\omega t}$.

∞ when $\omega \rightarrow m$. As a result, there could be a phase transition even in one dimension.

If for small σ , $\varepsilon(\sigma) = g\sigma^4 + g'\sigma^6 + \dots$ with $g < 0$, then $\sigma_1 = 0$ in equation (3). In this case, solitons exist for any particle number N , up to a maximum value $\bar{N}$ (corresponding to $\omega = \bar{\omega}$); it can be shown⁹ that because $dM/dN = \omega < m$, the soliton energy $M(N)$ is always less than Nm , insuring its stability. Thus, at low temperature, the thermodynamics of the system becomes dominated by the soliton phase, which resembles that of a Bose gas of rest mass $M(N)$, where N refers to the average particle number $N(T)$ varying with T . Because $M(N_1 + N_2) < M(N_1) + M(N_2)$ for any N_1 and N_2 , we have at $T = 0$, $N = \bar{N}$ (depending on the model, $\bar{N}$ can be quite large). As T increases, $N(T)$ decreases, and it becomes of order unity near T_c ; correspondingly, ω increases from $\bar{\omega}$ to m , accompanied by the related change of correlation length from $(m^2 - \bar{\omega}^2)^{-1/2}$ to ∞ . For $D = 1$, the specific heat is proportional to $T^{1/2+n}$ where $n > 0$. The value of n depends on the nonlinear interaction ε , and is associated with the gradual change of $N(T)$.

For $D > 1$, the radial part of ∇^2 is $\partial^2/\partial r^2 + (D-1)r^{-1}\partial/\partial r$; the latter gives a dissipative 'friction' to the corresponding 'energy' expression in the mechanical analogue given by equation (4). This sets a lower limit on the magnitude of the negative part of ε . In order to have solitons, ε cannot be arbitrarily small. Furthermore, stable solitons (that is, for $M(N) < Nm$) now exist only within a relatively limited range of N between, say, $\bar{N} - \Delta$ and $\bar{N}$. Again, the low-lying states are those of solitons, not particles of mass m . The corresponding specific heat is proportional to $T^{D/2+n}$. If Δ is narrow, then n is positive but small; this leads to, when $D = 2$, a nearly linear dependence on T for the specific heat.

A system of these solitons, like that of any bosons, can undergo Bose-Einstein condensation and thereby may exhibit superconductivity, in accordance with the work of M. R. Schafroth¹⁰.

In view of the rapid accumulation of new experimental results in this field and the apparent inadequacy of our present theoretic

cal understanding, exploration of different theoretical avenues may be necessary. It is in this spirit that this note is written.

I thank J. M. Luttinger for helpful discussions. This research was supported in part by the US Department of Energy.

Received 30 October; accepted 12 November, 1987.

1. Bednorz, J. G. & Müller, K. A. *Z. Phys.* **B64**, 189-193 (1986).
2. Bardeen, J., Cooper, L. N. & Schrieffer, J. R. *Phys. Rev.* **108**, 1175-1204 (1957).
3. Alexandrov, A. S., Ranninger, J. & Robaszkiewicz, S. *Phys. Rev. Lett.* **56**, 949-952 (1987).
4. Teller, E. *Proc. 1987 Erice Summer School* (ed. Zichichi, A.) (Academic, New York, in the press).
5. Anderson, P. W. *Science* **235**, 1196-1198 (1987).
6. Lindqvist, S., Tosatti, E., Tosi, M. & Yu Lu (eds) *Proc. Adriatico Res. Conf. High Temperature Superconductors* (World Scientific, Singapore, 1987).
7. Gan, Z. Z., Cu, G. J., Yang, G. Z. & Yang, Q. S. (eds) *Proc. Beijing Int. Workshop High Temperature Superconductivity* (World Scientific, Singapore, 1987).
8. Ginzburg, V. L. & Landau, L. D. *Zh. eksp. teor. Fiz.* **20**, 1064-1082 (1950).
9. Lee, T. D. *Particle Physics and Introduction to Field Theory*, Ch. 6 (Harwood, London, 1981).
10. Schafroth, M. R. *Phys. Rev.* **100**, 463-475 (1955).

Epitaxy in simple classical fluids in micropores and near-solid surfaces

C. L. Rhykerd Jr, M. Schoen*, D. J. Diestler & J. H. Cushman

Departments of Agronomy and Chemistry, Purdue University, West Lafayette, Indiana 47907, USA

* Present address: Institut für Naturwissenschaften, Lehrstuhl für Experimentalphysik, Universität Witten/Herdecke, Stockumer Strasse 10, 5810 Witten, FRG

The behaviour of fluids near solid-fluid interfaces is important in a number of fields¹⁻⁴ and is receiving renewed attention⁵. Previous computational studies⁶⁻¹¹ of the prototype split-pore—an atomic fluid contained between two infinite, plane-parallel solid surfaces (walls)—demonstrate that the fluid piles up in layers parallel to the walls. Yet none of these investigations finds long-range order within the fluid layers in directions parallel to the walls. In contrast, our Monte Carlo study of the prototype with structured walls reveals that at critical inter-wall separations the fluid layers take on solid-like order, that is, the fluid 'freezes' to a solid reflecting the structure of the wall. This epitaxy decreases with increasing inter-wall separation, but persists indefinitely in the contact layer. Such epitaxy may occur at aqueous-solid interfaces in geological^{1,2,4} and biological^{2,3} systems.

In our model each wall consists of a square 'unit cell' of the (100) plane of a face-centred-cubic (f.c.c.) lattice (Fig. 1). One wall lies in the $z = 0$ plane and the other in the $z = h$ plane. The

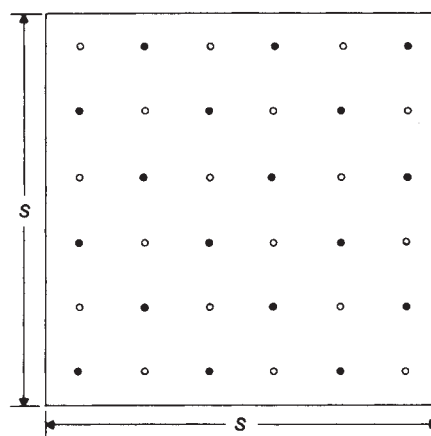


Fig. 1 Square 'unit cell' of the (100) plane of the f.c.c. lattice. The filled circles represent atoms in the $z = 0$ plane; the open circles refer to the $z = h$ plane, shown out of registry. When the walls are in registry, the filled and open circles coincide.

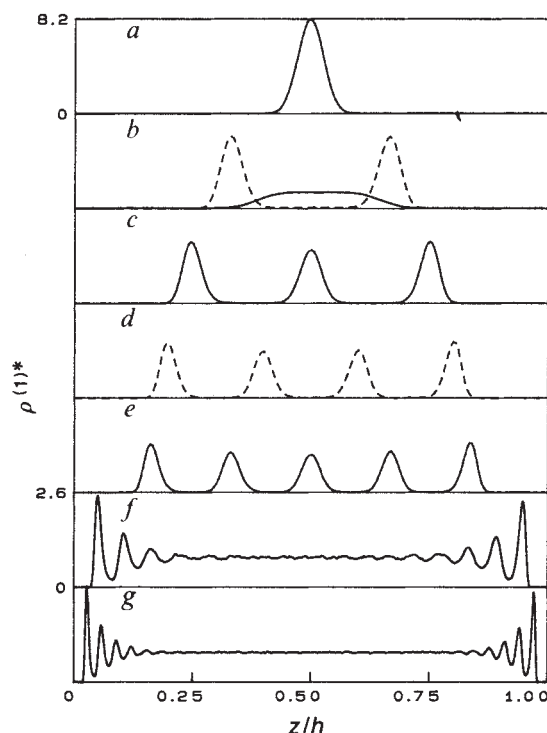


Fig. 2 Local density profiles as a function of inter-wall separation h^* for pore fluid at a temperature T^* of 1.0. The surface density of wall atoms is $N_s/s^{*2} = 0.7827$, where $N_s = 50$, the number of wall atoms. Panel *a*, $h^* = 1.37$, $\mu^* = -9.00$; *b*, $h^* = 2.20$, $\mu^* = -9.26$; *c*, $h^* = 3.05$, $\mu^* = -9.60$; *d*, $h^* = 3.90$, $\mu^* = -9.83$; *e*, $h^* = 4.90$, $\mu^* = -10.0$; *f*, $h^* = 16.5$, $\mu^* = -10.29$; *g*, $h^* = 30.0$, $\mu^* = -10.29$. In panels *a-e* the scale of $\rho^{(1)*}$ ranges from 0 to 8.2; the solid and dashed curves respectively correspond to walls in and out of registry. In panels *f* and *g* the scale of $\rho^{(1)*}$ ranges from 0 to 2.6; the walls are out of registry.

walls are positioned in the x - and y -direction so that they are either adjacent (out of registry, as shown in Fig. 1) or alternate (in registry) layers of the (100) face of the f.c.c. lattice. The space between the walls is filled with rare-gas atoms identical to the wall atoms. The total potential energy is a sum of pairwise Lennard-Jones (12, 6) interactions

$$u(r) = 4\epsilon[(\sigma/r)^{12} - (\sigma/r)^6]$$

where r is the distance between the atoms of a pair. To represent a slit-pore of infinite extent, we impose periodic boundary conditions on the 'unit cell' in the x - and y -directions.

The chemical potential of the pore fluid should be fixed, in order to relate the computational results to key experiments. Consequently we use the grand canonical ensemble Monte Carlo method¹². The thermodynamic state of the pore fluid is specified by the fixed chemical potential μ , temperature T and volume $V = s^2h$, where s is the side of the unit cell. Equilibrium properties are computed as averages over fluid-atom configurations generated by a Markov process so as to be representative of a grand canonical ensemble¹³. The initial configuration is random and, depending on the thermodynamic state, 1 million to 15 million configurations (steps) are sampled. The configurational-energy change is corrected at each step for long-range interactions. The side s of the unit cell is taken sufficiently large that the results are independent of s . This is assured by performing the calculations for a sequence of unit cells of increasing size.

The structure of the pore fluid is manifested by the singlet distribution function $\rho^{(1)}(\mathbf{r})$, or local density, and the pair distribution function $\rho^{(2)}(\mathbf{r}_1, \mathbf{r}_2)$. On account of the symmetry of the

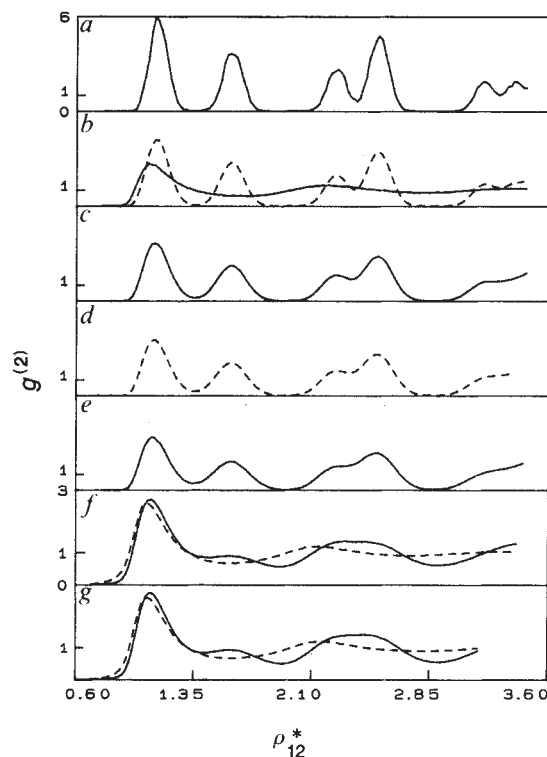


Fig. 3 In-plane pair correlation functions corresponding to the density profiles of Fig. 2. In panels *a-e*, only $g^{(2)}$ for the contact layer is shown, as $g^{(2)}$ values for the inner layers are indistinguishable from it; the scale of $g^{(2)}$ ranges from 0 to 6.0. In panels *f* and *g*, solid and dashed curves correspond respectively to contact and first inner layers; the scale of $g^{(2)}$ ranges from 0 to 3.0.

pore, $\rho^{(1)}$ depends (approximately) only on z . It is expressed as

$$\rho^{(1)}(z) = \langle N(z) \rangle / s^2 \Delta z \quad (1)$$

where $N(z)$ is the number of atoms in a layer of fluid of thickness $\Delta z = 0.01\sigma$ and the brackets signify the average over configurations. The pair distribution function, which is proportional to the probability of finding a (reference) molecule in d^3r_1 and simultaneously another molecule in d^3r_2 , is related to the local density through the pair correlation function $g^{(2)}(\mathbf{r}_1, \mathbf{r}_2)$ by

$$\rho^{(2)}(\mathbf{r}_1, \mathbf{r}_2) = \rho^{(1)}(\mathbf{r}_1)\rho^{(1)}(\mathbf{r}_2)g^{(2)}(\mathbf{r}_1, \mathbf{r}_2) \quad (2)$$

By symmetry $g^{(2)}$ depends only on z_1 and the cylindrical coordinates ρ_{12} and z_{12} of atom 2 with respect to atom 1. We restrict our attention to the in-plane ($z_{12} = z_2 - z_1 = 0$) pair correlation function, given by

$$g^{(2)}(z_1, \rho_{12}) = \langle N(z_1, \rho_{12}) \rangle / 2\pi\rho_{12}\Delta\rho_{12}\Delta z_{12}\rho^{(1)}(z_1) \quad (3)$$

where $\langle N(z_1, \rho_{12}) \rangle$ is the average number of atoms lying in an annulus of radius ρ_{12} centred on reference atom 1. The thickness of the annulus is $\Delta\rho_{12} = 0.02\sigma$; its height Δz_{12} is determined by $\rho^{(1)}$ as described below.

Figure 2 shows the dependence of the local density profile on the separation $h^* = h/\sigma$ between the walls. Distance and energy are expressed in units of σ and ϵ , respectively, so that the reduced density is $\rho^* = \rho\sigma^3$, the reduced chemical potential is $\mu^* = \mu/\epsilon$ and the reduced temperature is $T^* = k_B T/\epsilon$, where k_B is Boltzmann's constant. As expected from previous work⁶⁻¹¹ the pore fluid orders itself in layers (peaks in $\rho^{(1)}$) parallel to the wall.

Figure 3 displays in-plane pair correlation functions for 'imaginary' layers of fluid determined by $\rho^{(1)}$. The layers are centred on the peaks in $\rho^{(1)}$ and bounded by planes intersecting

the nearest minima in $\rho^{(1)}$. In equation (3) Δz_{12} is the distance between the bounding planes and $\rho^{(1)}(z_1)$ is replaced by its average over Δz_{12} .

For the smallest distance between walls in registry there is a single sharp layer with a $g^{(2)}$ characteristic of a solid phase. Indeed the positions and heights of the peaks in $g^{(2)}$ (Fig. 3a) agree with those of the pair correlation function for the (100) plane of the f.c.c. lattice. As h^* increases the in-plane order decreases until $g^{(2)}$ becomes typical of a liquid (Fig. 3b) at $h^* = 2.20$. However, if the walls are moved out of registry with h^* held fixed, two sharp highly ordered layers appear. With the walls in registry 'phase transitions' at certain precise inter-wall separations in the range $1.37 < h^* < 5.2$ produce one, three or five solid-like layers (Fig. 2a, c, e). Likewise, with the walls out of registry, phase transitions yield two or four solid layers (Fig. 2b, d). As the number of layers increases, the in-plane order decreases (Fig. 3a-e), approaching that of an amorphous solid (Fig. 3e). Yet vestiges of in-plane order are apparent in the contact layer even for the largest inter-wall separations examined (Fig. 3f, g).

The pore fluid 'freezes' into an f.c.c.-like lattice at certain critical inter-wall separations less than 6 atomic diameters. One, three and five fluid layers would stack between two alternate (in registry) f.c.c. solid layers whereas two and four layers would stack between adjacent (out of registry) f.c.c. solid layers. The

transitions, however, do not occur at precisely the right inter-wall separations corresponding to an f.c.c. solid having the density of the solid walls. Moreover, as the pore fluid is in equilibrium with bulk fluid in a thermodynamic state within the liquid region of the Lennard-Jones (12, 6) phase diagram¹⁴ it cannot form a true f.c.c. solid phase.

We are grateful to the US Department of Energy for financial support and to Mr Jin-Song Chen for preparation of Figs 2 and 3.

Received 22 June; accepted 22 September 1987.

1. Forslind, E. & Jacobsson, A. in *Water: A Comprehensive Treatise*, Vol. 5 (ed. Franks, F.) 75-132 (Plenum, New York, 1975).
2. Clifford, J. in *Water: A Comprehensive Treatise*, Vol. 5 (ed. Franks, F.) 173-248 (Plenum, New York, 1975).
3. Etzler, F. M. & Drost-Hansen, W. *Adv. Chem.* **188**, 486-497 (1980).
4. Low, P. F. *Langmuir* **3**, 18-25 (1987).
5. Robinson, A. L. *Science* **23**, 150-151 (1987).
6. Rowley, L. A., Nicholson, D. & Parsonage, N. G. *Molec. Phys.* **31**, 365-387; 389-407 (1976).
7. Lane, J. E. & Spurling, T. H. *Chem. Phys. Lett.* **67**, 107-108 (1979); *Aust. J. Chem.* **29**, 2103-2121 (1976).
8. Abraham, F. F. *J. chem. Phys.* **68**, 3713-3716 (1979).
9. Snook, I. K. & van Megen, W. *J. chem. Phys.* **72**, 2907-2913 (1980).
10. Toxvaerd, S. *J. chem. Phys.* **74**, 1998-2005 (1981).
11. Magda, J. J., Tirrell, M. & Davis, H. T. *J. chem. Phys.* **83**, 1888-1901 (1985).
12. Adams, D. J. *Molec. Phys.* **29**, 307-311 (1975).
13. Valleau, J. P. & Torrie, G. M. in *Modern Theoretical Chemistry*, Vol. 5 (ed. Berne, B. J.) 169-194 (Plenum, New York, 1977).
14. Hansen, J.-P. & Verlet, L. *Phys. Rev.* **184**, 151-161 (1969).

Precision spectroscopy of hydrogen and deuterium

M. G. Boshier*, P. E. G. Baird*, C. J. Foot*,
E. A. Hinds†, M. D. Plimmer*, D. N. Stacey*,
J. B. Swan‡, D. A. Tate§, D. M. Warrington||
& G. K. Woodgate*

* Clarendon Laboratory, Parks Road, Oxford OX1 3PU, UK

† Physics Department, Yale University, New Haven, Connecticut 06520, USA

‡ Department of Physics, University of Western Australia, Nedlands, Western Australia 6009, Australia

§ Department of Physical Chemistry, Lensfield Road, Cambridge CB2 1EP, UK

|| Physics Department, University of Otago, PO Box 56, Dunedin, New Zealand

The hydrogen atom is an important testing ground for fundamental physical theory, because it is the simplest stable atomic system and its properties can be calculated to enormous precision. Measurements on hydrogen played a major role in the development of quantum electrodynamics (QED), a theory of electromagnetic interactions which includes quantization of the radiation field^{1,2}. Tests of QED in hydrogen involve precision measurements of transition frequencies; the transition generally studied is between two excited states ($2s^2S_{1/2}$ and $2p^2P_{1/2}$). Much higher precision is possible if a transition involving the ground state is studied, but this potential has not so far been exploited because of experimental difficulties. We have now carried out an experiment of this type which has given preliminary results and opens new possibilities for precision tests of QED in the future.

The most recent experimental results^{3,4} for the interval $2s^2S_{1/2}-2p^2P_{1/2}$, which is about 1 GHz and almost entirely due to QED effects, have quoted uncertainties of the order of 10 kHz. It would be of great interest to improve this precision, but the problem is the 100 MHz natural width of the $2p^2P_{1/2}$ level; the centre of the resonance is already being located to about 1 part in 10^4 of its width. For this reason, increasing attention has been given to the two-photon transition $1s^2S_{1/2}-2s^2S_{1/2}$, which has an intrinsic width of only 1.3 Hz. This transition can be studied

without Doppler broadening by laser spectroscopy⁵, so that in principle it offers much the better long-term prospect for testing QED. In contrast to the $2s-2p$ interval, however, QED effects contribute only a tiny fraction of the $1s-2s$ interval—about 7 GHz out of 2.5×10^6 GHz. There is thus the difficulty of extracting the small quantity of interest from the experimental measurement.

The problem can be overcome by using a scheme first realized at Stanford⁶, in which the radiation at 243 nm required for the Doppler-free two-photon excitation is produced by frequency doubling laser light at 486 nm. Since the $n=2-n=4$ (H β) transition in hydrogen is also at 486 nm, the small frequency difference between one quarter of the $1s-2s$ interval and a component of H β can be directly measured. The QED displacement (or 'Lamb shift') of the $1s$ level is a significant fraction of this interval. The method thus avoids the need for either a precisely-known external frequency standard or an accurate value of the Rydberg constant. Frequency doubling at 486 nm is the important feature of the scheme, and has formed the basis of several subsequent studies of the $1s-2s$ transition⁷⁻¹⁰; however, the use of pulsed rather than continuous-wave (c.w.) lasers has always been a major limitation. Although pulsed lasers make the experiment much easier, they do not allow the potential of the method to be fully exploited because of their large linewidths and the distorting effects of the laser amplifiers. They also do not permit the use of accurate heterodyne techniques to measure frequency intervals. To avoid these problems, the group at Stanford has carried out a c.w. experiment in which the 243-nm radiation was produced by sum-frequency mixing^{11,12} but this technique does not achieve the important goal of providing a precise link between the frequencies of two different hydrogen transitions.

We have now observed and calibrated the two-photon signal with c.w. radiation, generated by frequency doubling. So far, we have calibrated the transition with respect to an external frequency standard; however, the results are already the best $1s$ Lamb shift measurements available for hydrogen and deuterium, and the way is now open for further substantial improvements in precision using the H β comparison scheme outlined above.

A diagram of the experimental arrangement is shown in Fig. 1. The 243-nm radiation was produced in a crystal of β -barium borate placed within the cavity of a ring dye laser operating single mode at 486 nm with a bandwidth of ~ 1 MHz. Although our first observation of the two-photon signal was with a crystal

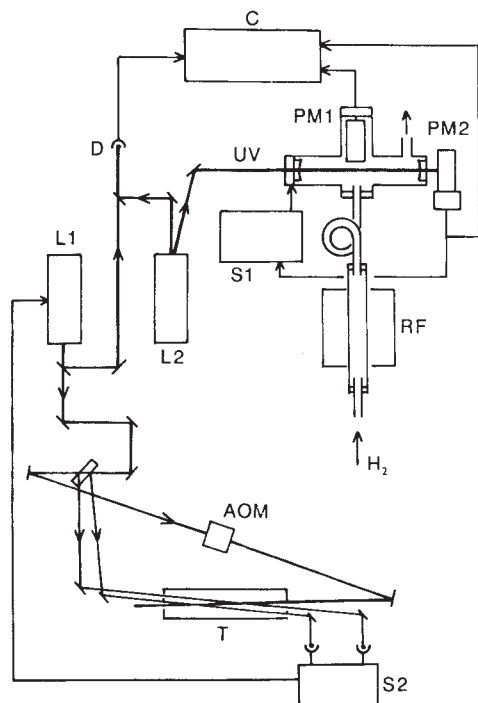


Fig. 1 Diagram of the apparatus. L1, L2, tunable dye lasers; UV, ultraviolet radiation (243 nm); RF, radio-frequency dissociation of flowing molecular hydrogen; PM1, signal photomultiplier (Lyman- α detector). PM2, photomultiplier for cavity locking and signal normalization; S1, cavity length servo-control; C, computer; AOM, acousto-optic modulator; T, heated quartz cell containing tellurium; S2, laser frequency servo-control; D, fast photodiode.

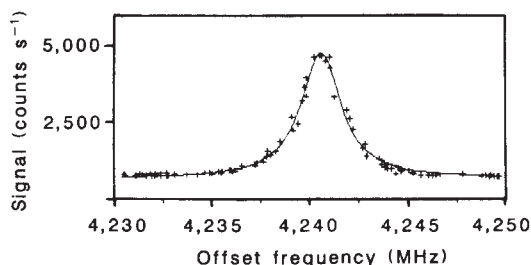


Fig. 2 The stronger component of the $1s-2s$ two photon transition in deuterium. The signal is the normalized Lyman- α fluorescence observed as a function of the frequency difference between lasers L1 and L2 (Fig. 1) when L1 is locked to the b_2 transition in $^{130}\text{Te}_2$. The measured offset frequency is 20 MHz greater than the true value because of the shift introduced by the acousto-optic modulator. The pressure in the deuterium cell was 270 mtorr.

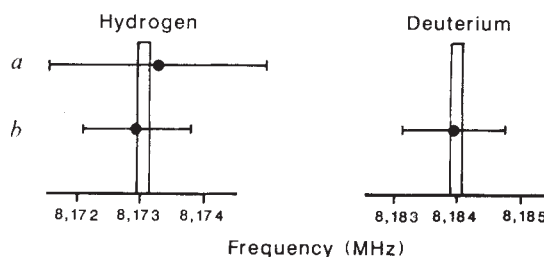


Fig. 3 Comparison of measured $1s$ Lamb shifts with theory², the limits of which are shown as vertical lines. *a*, From ref 12; *b*, present work. No earlier work of comparable precision exists for deuterium.

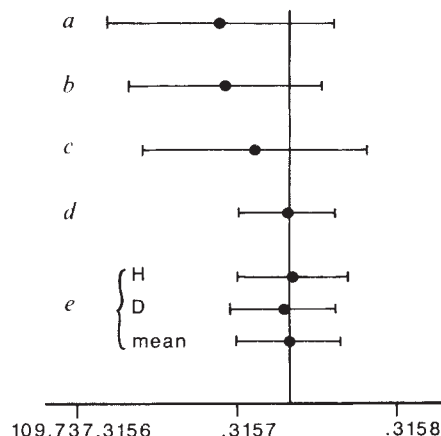


Fig. 4 The present results for the Rydberg constant compared with other recent values. *a*, From ref 19; *b*, ref. 20; *c*, ref. 12; *d*, ref. 15; *e*, present work. The vertical line is drawn through the mean of the new results.

A second tunable dye laser, also operating at 486 nm, was locked to a transition in $^{130}\text{Te}_2$ by Doppler-free saturation spectroscopy. An acousto-optic modulator was used as a light chopper; this technique shifted the resonance by precisely 20 MHz and introduced negligible error. Some of the laser output was mixed with the fundamental from the frequency doubling laser on a fast photodiode, thus providing a heterodyne signal. Different tellurium transitions (designated b_2 and b_1 respectively¹⁴) were used for our measurements on hydrogen and deuterium; both had themselves been calibrated at the National Physical Laboratory in terms of the ionide-stabilized helium-neon standard¹⁴. For both gases, measurements were made at various cell pressures in the range 20–300 mtorr to permit an extrapolation to zero pressure.

This extrapolation gives 1,379.53(7) MHz for hydrogen and 4,220.77(7) MHz for deuterium, the tellurium transition having the lower frequency in each case. The figure in parenthesis is the uncertainty in the last quoted digit, and corresponds to one standard deviation. Allowing for the precisely known hyperfine structures, we find frequency separations

$$f(1s-2s)/4 - f(\text{Te}) = 1,457.21(11) \text{ MHz (H)}, \\ 4,244.64(8) \text{ MHz (D)}$$

where the errors now take into account instrumental effects which can cause shifts in the measured positions of the tellurium lines. The absolute frequencies are found to be

$$f(1s-2s, \text{H}) = 2,466,061,414.13(79) \text{ MHz} \\ f(1s-2s, \text{D}) = 2,466,732,408.55(73) \text{ MHz}$$

The uncertainty in the tellurium frequency standards gives a contribution of 0.66 MHz to the error in each case. We have

of urea, β -barium borate is superior in almost all respects¹³. The ultraviolet output, typically 2 mW, was enhanced by more than an order of magnitude in a servo-controlled optical cavity (inside the hydrogen cell). Hydrogen molecules were dissociated in a radio-frequency discharge, the atoms then being transported to the cell in a steady flow system through a short loop of Teflon tubing. This material has a low surface recombination rate, and the loop prevents discharge light from reaching the interaction region. The laser was scanned through the position of the two-photon resonance; the excitation was monitored by observing Lyman- α radiation, emitted by atoms collisionally transferred from $2s$ to $2p$ states. For both isotopes, the hyperfine structure gives rise to two components, and the stronger component was measured in each case; a typical recording for deuterium is shown in Fig. 2.

taken into account unpublished comparisons of iodine-stabilized lasers used as frequency standards at the National Physical Laboratory and the National Bureau of Standards.

To obtain the Lamb shifts of the $1s$ levels we use the recent determination of the Rydberg constant¹⁵ and the known $2s$ and $2p$ Lamb shifts. Our definition of the Lamb shift¹⁶ is the sum of all QED corrections plus the finite nuclear size correction; we obtain 8,172.93(84) MHz for hydrogen and 8,183.96(80) MHz for deuterium, in excellent agreement with theory (Fig. 3). The latter result has an uncertainty 35 times smaller than the previous best value.

As we have stressed, the significance of this experiment lies in the promise of much higher accuracy when the technique is applied to the direct comparison of hydrogen wavelengths; the next step is therefore to replace the tellurium spectrometer with a beam of metastable ($2s$) hydrogen atoms. We can then compare directly the $1s-2s$ and $2s-4s$ transitions, the latter being preferred to a $2s-4p$ transition because of its narrower linewidth; it can be excited in the presence of a weak electric field. At present the fractional precision of the $n=2$ Lamb shift measurements is still an order of magnitude higher than that of the $n=1$ results; however, the new data do confirm the predicted state dependence of the QED terms. An alternative interpretation of our results is possible if it is supposed that the theoretical values of the $1s$ Lamb shifts are correct to significantly better than our present uncertainty. Our data then give values of the Rydberg constant as follows: $R = 109,737.315\,735(35)\text{ cm}^{-1}$ from hydrogen and $109,737.315\,729(33)\text{ cm}^{-1}$ from deuterium. These are compared with other recent measurements in Fig. 4. Finally, we obtain a value for the H-D $1s-2s$ isotope shift of 670,994.4(9) MHz, to be compared with 670,994.48(10) MHz from theory². This experimental result can be improved by more than an order of magnitude by the use of wide-band heterodyne techniques¹⁷, so avoiding the need for frequency standards. It will then provide a useful consistency check of the electron-proton mass ratio as given by ion-trap experiments¹⁸.

We are glad to acknowledge the technical assistance from the Clarendon Laboratory, Oxford, in particular by G. Read, S. Giles, S. Smith and C. W. Goodwin. We are grateful to A. I. Ferguson for the loan of the calibrated tellurium cell. The work was supported by the SERC, which also awarded studentships to M.D.P. and D.A.T., a Visiting Fellowship to E.A.H. and a Senior Fellowship to D.N.S.. M.G.B. was supported by the Royal Commission for the Exhibition of 1851 and Christ Church, Oxford. C.J.F. holds a Royal Society Research Fellowship.

A short report of this work was presented at the Eighth International Conference on Laser Spectroscopy held at Are, 22–26 June 1987.

Received 28 July; accepted 7 October 1987.

1. Ermolaev, A. M. in *Progress in Atomic Spectroscopy* Part A (eds Hanle, W. & Kleinpoppen, H.) 149–182 (Plenum, New York, 1978).
2. Erickson, G. W. *J. Phys. Chem. Ref. Data* **6**, 831–869 (1977).
3. Lundeen, S. R. & Pipkin, F. M. *Phys. Rev. Lett.* **46**, 232–235 (1981).
4. Cosens, B. L. *Phys. Rev.* **173**, 49–55 (1968).
5. Baklanov, E. V. & Chebotayev, V. P. *Opt. Commun.* **12**, 312–314 (1974).
6. Hänsch, T. W., Lee, S. A., Wallenstein, R. & Wieman, C. *Phys. Rev. Lett.* **34**, 307–309 (1975).
7. Lee, S. A., Wallenstein, R. & Hänsch, T. W. *Phys. Rev. Lett.* **35**, 1262–1266 (1975).
8. Wieman, C. & Hänsch, T. W. *Phys. Rev. A* **22**, 192–205 (1980).
9. Hildum, E. A., Boesl, U., McIntyre, D. H., Beausoleil, R. G. & Hänsch, T. W. *Phys. Rev. Lett.* **56**, 576–579 (1986).
10. Barr, J. R. M., Girkin, J. M., Tolchard, J. M. & Ferguson, A. I. *Phys. Rev. Lett.* **56**, 580–583 (1986).
11. Foot, C. J., Couillaud, B., Beausoleil, R. G. & Hänsch, T. W. *Phys. Rev. Lett.* **54**, 1913–1916 (1985).
12. Beausoleil, R. G. *et al. Phys. Rev. A* **35**, 4878–4881 (1987).
13. Kato, K. *IEEE J. Quant. Electr.* **QE-22**, 1013–1014 (1986).
14. Barr, J. R. M. *et al. Opt. Commun.* **65**, 217–221 (1985).
15. Zhao, P., Lichten, W., Layer, H. & Berquist, J. *Phys. Rev. Lett.* **58**, 1293–1295 (1987).
16. Johnson, W. R. & Soff, G. *At. Data Nucl. Data Tables* **33**, 405–446 (1985).
17. Burghardt, R., Hoeffgen, H., Meisel, G., Reinert, W. & Vowinkel, B. *Appl. Phys. Lett.* **35**, 498–500 (1979).
18. van Dyck, R. S. Jr, Moore, F. L., Farnham, D. L. & Schwinberg, P. B. *Bull. Am. Phys. Soc.* **31**, 224 (1986).
19. Zhao, P., Lichten, W., Layer, H. P. & Berquist, J. C. *Phys. Rev. A* **34**, 5138–5141 (1986).
20. Biraben, F., Garreau, J. C. & Julien, L. *Europhys. Lett.* **2**, 925–932 (1986).

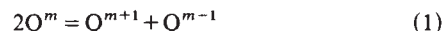
Identification of multiple structural species in silicate glasses by ^{29}Si NMR

Jonathan F. Stebbins

Department of Geology, Stanford University, Stanford, California 94305, USA

Knowledge of the structure of silicate glasses is one key to understanding their chemical and physical properties, as well as those of the high-temperature liquids from which they form. Because first-neighbour ordering into SiO_4 tetrahedra is well known, much of the recent discussion about glass structure has concerned intermediate-range order. One of the most fundamental questions that can be posed is the distribution of silicate tetrahedra with n bridging oxygens (shared between two tetrahedra), labelled as ' Q^n ' ($0 \leq n \leq 4$). This distribution is probably important in determining the properties of melts and glasses, but considerable controversy exists concerning the identification and quantification of such species. Here I present new evidence from ^{29}Si NMR (nuclear magnetic resonance) spectroscopy that clearly resolves one significant part of this dispute: at least in sodium silicate glasses, a variety of species is present that is greater than that required by composition alone.

For a stoichiometric composition such as $\text{Na}_2\text{Si}_2\text{O}_5$, the entire glass structure could be made up of a single Q species, as in the crystal (Q^3 for this composition). For binary non-stoichiometric compositions, no more than two species are required. The central question is whether a range of species beyond that required by composition alone is actually present in the glass. In particular, disproportionation reactions such as:



($0 < m < 4$) have often been called on to produce multiple species^{1–4}. Relatively small populations of species could be of great energetic significance, particularly in controlling component activities and configurational entropies of glasses and melts.

Distributions of species (including analogous models that include free oxide ions) have played fundamental parts in thermodynamic models of melt chemistry^{5–8}, but the direct evidence for this distribution is somewhat uncertain. Infrared absorption and Raman spectra have been interpreted to indicate that multiple species are present in alkali silicate glasses^{9–11}. On the other hand, the results of several X-ray diffraction studies of similar systems have been interpreted as indicating the presence of single Q species^{12,13}.

^{29}Si magic-angle-sample spinning (MAS) NMR has been used in a number of studies in attempts to quantify such species distributions. Interpretations of MAS spectra of alkali silicate glasses have differed, however. In several cases, multiple gaussian peaks have been fitted to spectra to obtain distributions of three, four or five species^{1,14,15}. In other cases, peaks were fitted only to clearly distinguishable features in the spectra^{16–20}. In most cases, with the latter assumption, only one or two Q^n species were reported, as determined by stoichiometry alone. Earlier studies have shown that much of this controversy is due to the difficulty in obtaining unique fits to often closely overlapping peaks in MAS spectra²¹.

Figure 1a shows ^{29}Si NMR spectra, taken on non-spinning samples, of three sodium silicate glasses with analysed sodium contents of 34.7, 37.6 and 41.7 ± 0.5 wt%, corresponding to 34.2, 37.0 and 41.1 mol% Na_2O . Experimental conditions (listed in Fig. 1 legend) were chosen to allow quantitative analysis: in particular, the delay between pulses was long enough such that all parts of the spectra were fully relaxed. In these 'static' spectra, two components are clearly visible. The major part of each is a typical powder pattern for a site with near uniaxial symmetry, and is quite similar to the static spectrum for crystalline

$\text{Na}_2\text{Si}_2\text{O}_5$, with some broadening due to positional disorder. This broad component of the static spectrum is due to a mixture of Q^3 with some (inseparable) Q^2 species.

The most novel features of the static spectra are the small, clearly distinguishable, narrow peaks superimposed on this major component and centered at ~ -100 p.p.m. Based on shape and chemical shift, these can only be attributed to Q^4 species²². These are accentuated in the static spectra because of the small magnitude of the chemical shift anisotropy (CSA) of this symmetrical site type. Data on pure SiO_2 glass¹ indicates that the Q^4 MAS linewidth is probably at least 10–15 p.p.m. because of bond angle disorder, a value that is not much less than the 20 p.p.m. linewidths of the Q^4 peaks fitted to the static spectra shown here. This small change from MAS to static spectrum, in contrast to the large change for the less symmetrical sites, allows the minority Q^4 species to be counted.

The areas of the Q^4 peaks are shown by cross-hatching in Fig. 1a; values of these areas were determined by simulation of the spectra with lineshapes derived from end-member compositions and the approximation that the sum of the mole fractions of Q^4 , Q^3 , and Q^2 is equal to 1. Abundances of Q^4 in three glasses are 7.2%, 4.0% and $2.0\% \pm 1\%$ absolute. The predominant Q^3 pattern was simulated by a broadened CSA tensor with $\sigma_{11} = -48 \pm 5$, $\sigma_{22} = -55 \pm 3$ and $\sigma_{33} = -162 \pm 5$ p.p.m. The isotropic chemical shift calculated from these results is -88.3 p.p.m., in good agreement with the value from the MAS spectra (Fig. 1b) of about -88.5 p.p.m. From data for Na_2SiO_3 glass and crystal, a broadened CSA tensor with $\sigma_{11} = -18 \pm 10$, $\sigma_{22} = -61 \pm 3$ and $\sigma_{33} = -147 \pm 10$ p.p.m. was chosen for the Q^2 species, and an area appropriate for the composition and the observed Q^4 abundance was included in the simulation.

All three compositions are outside the range where liquid-liquid phase separation occurs, and are shown by the MAS spectra (Fig. 1b) to be crystal-free. Repeated grinding and remelting produced no significant changes in the spectra, indicating that the samples represent the structure of homogeneous liquids. In all three cases, stoichiometry alone would predict that mixtures of only Q^3 and Q^2 species should be present.

The presence of the Q^4 population could be inferred from the MAS spectra alone only by uncertain curve-fitting of the 'tail' of the main Q^3 peak. The MAS data do, however, clearly reveal the presence of the expected levels of Q^2 species as distinct peaks. Perhaps coincidentally, the measured Q^4 abundance in the 34 mol% Na_2O glass is the same as that estimated from our earlier MAS work on a similar material¹.

Data for the Q^4 abundances in the 34% glass allows the estimation of 'equilibrium constant' $K_3 = 0.011 \pm 0.005$ for reaction (1) with $m = 3$, assuming that the $m = 2$ and $m = 1$ reactions are insignificant, that activity coefficients are 1 and constraints of mass balance. K_2 cannot be precisely estimated, but also appears to be non-zero. The observed species distribution is considerably more ordered than a completely random mixing of bridging and non-bridging oxygens²³, but is more disordered than a 'stoichiometric' speciation. The distribution is narrower than that estimated from MAS spectra in the $\text{Li}_2\text{O}-\text{SiO}_2$ system¹⁵ and preliminary work in this laboratory on the static spectra in that binary. This is in agreement with trends predicted from MAS NMR and other spectroscopies, thermodynamics and molecular orbital calculations^{1,4,11,24}. A decrease in the disproportionation is expected in potassium silicate glasses, explaining why only two species were observed in an MAS NMR study of this system¹⁹.

The ideal entropy of mixing of the estimated Q species distribution for $\text{Na}_2\text{Si}_2\text{O}_5$ glass is $\sim 10 \text{ J K}^{-1} \text{ mol}^{-1}$. Although this is a very crude approximation, it does indicate that approximately half the configurational entropy of this glass at its glass transition²⁵ could be due to 'speciation' reactions. Positional disorder, including bond angle and distance variations, must also play a major part.

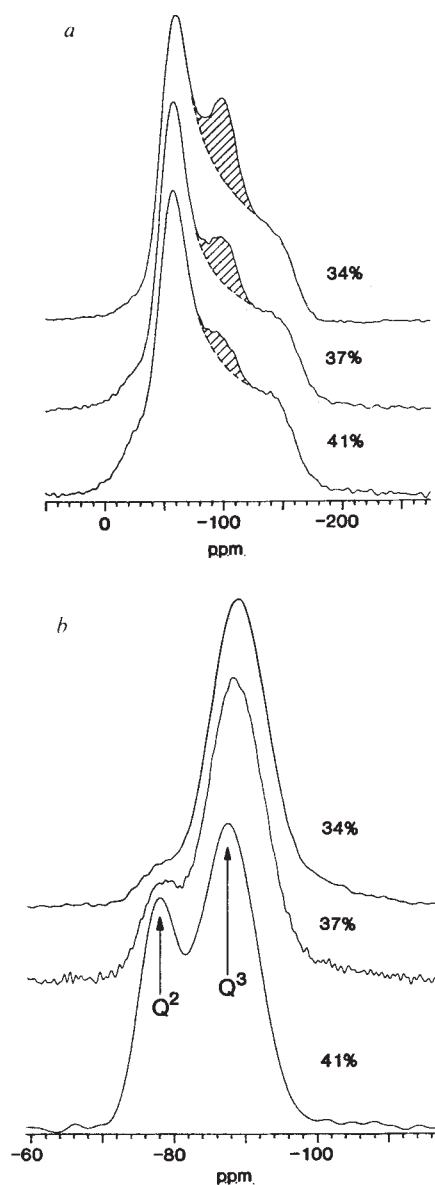


Fig. 1 ^{29}Si NMR spectra, of sodium silicate glasses with mol% Na_2O as shown: a, taken without sample spinning; b, taken with magic-angle spinning (MAS) of samples. Cross-hatching shows areas of Q^4 peaks. Samples were enriched to 57% ^{29}Si . Data were collected with a modified Bruker HX-360 spectrometer at 71.5 MHz at the Stanford Magnetic Resonance Laboratory. In a, several hundred signal averages were collected for each spectrum, using a $4 \mu\text{s}$ (30°) pulse with 60 s delay between pulses, collecting 256 points and zero-filling to 2K points. No apodization was used. A minor amount of broadening due to Si-Si and Na-Si dipolar interactions is present. In b, approximately 100 signal averages with 512 points each were collected. Spinning rates of about 6,400 Hz kept spinning sidebands (not shown) to within a few per cent of the central peak areas.

Finally, preliminary work indicates that there is a slight increase in the abundance of Q^4 species in glasses with faster quench rates and therefore higher glass transition temperatures. K_3 will therefore be somewhat temperature-dependent. The role of this change in 'static' disorder in the energetics of the liquids remains to be assessed: the contribution of the dynamical effects observed in NMR on the melts themselves probably still dominates²⁶⁻²⁸.

This study was supported by an NSF grant and by NIH grants to the Stanford Magnetic Resonance Laboratory. I thank Drs O. Jardetsky and R. Domenick of SMRL for assistance.

Received 13 July; accepted 6 October 1987.

1. Murdoch, J. B., Stebbins, J. F. & Carmichael, I. S. E. *Am. Miner.* **70**, 332-343 (1985).
2. Mysen, B. O., Virgo, D. & Scarfe, C. M. *Am. Miner.* **65**, 690-710 (1980).
3. McMillan, P. *Am. Miner.* **69**, 645-659 (1984).
4. DeJong, B. H. W. S., Keefer, K. D., Brown, G. E., Jr & Taylor, C. M. *Geochim. cosmochim. Acta* **45**, 1291-1308 (1981).
5. Charles, R. J. *Phys. Chem. Glasses* **10**, 169-178 (1969).
6. Gaskell, D. R. *Metal. Trans.* **8B**, 131-145 (1977).
7. Masson, C. R. *J. noncryst. Solids* **25**, 3-41 (1977).
8. Hess, P. C. in *Physics of Magmatic Processes* (ed. Hargraves, R. B.) 3-44 (Princeton University Press, 1980).
9. Brawer, S. A. & White, W. B. *J. chem. Phys.* **63**, 2421-2432 (1975).
10. Sharma, S. K., Virgo, D. & Mysen, B. O. *Carnegie Inst. Wash. Year Book* **77**, 649-652 (1978).
11. Matson, D. W., Sharma, S. K. & Philpotts, J. A. *J. noncryst. Solids* **58**, 323-352 (1983).
12. Yasui, I., Hasegawa, H. & Imaoka, M. *Phys. Chem. Glasses* **24**, 65-71 (1983).
13. Imaoka, M., Hasegawa, H. & Yasui, I. *Phys. Chem. Glasses* **24**, 72-78 (1983).
14. Selvaraj, U., Rao, K. J., Rao, C. N. R., Klinowski, J. & Thomas, J. M. *Chem. Phys. Lett.* **114**, 24-27 (1985).
15. Schramm, C. M., de Jong, B. H. W. S. & Parziale, V. E. *J. Am. chem. Soc.* **106**, 4396-4402 (1984).
16. Grimmer, A.-R. *et al. Phys. Chem. Glasses* **25**, 105-109 (1984).
17. Dupree, R., Holland, D. & Williams, D. S. *J. noncryst. Solids* **68**, 399-410 (1984).
18. Dupree, R., Holland, D. & Williams, D. S. *J. Phys. Coll.* **C8**, 46, 119-123 (1985).
19. Grimmer, A.-R. & Müller, W. *Monatsh. Chem.* **117**, 799-803 (1986).
20. Rao, C. N. R. *et al. Angew. Chem. Int. edn, Eng.* **24**, 61-62 (1985).
21. Schneider, E., Stebbins, J. R. & Pines, A. *J. noncryst. Solids* **89**, 371-383 (1987).
22. Smith, K. A., Kirkpatrick, R. J., Oldfield, E. & Henderson, D. M. *Am. Miner.* **68**, 1206-1215 (1983).
23. Lacy, E. D. *Phys. Chem. Glasses* **6**, 171-180 (1965).
24. DeJong, B. H. W. S. & Brown, G. E. *Geochim. cosmochim. Acta* **44**, 1627-1642 (1980).
25. Richet, P. *Geochim. cosmochim. Acta* **48**, 471-483 (1984).
26. Schneider, E. thesis, California Univ. (1985).
27. Stebbins, J. F., Murdoch, J. B., Schneider, E., Carmichael, I. S. E. & Pines, A. *Nature* **314**, 250-252 (1985).
28. Liu, S.-B., Pines, A., Brandriss, M. & Stebbins, J. F. *Phys. chem. Miner.* (in the press).

Suppression of earthquakes by large continental ice sheets

Arch C. Johnston

Center for Earthquake Research, Memphis State University,
Memphis, Tennessee 38152, USA

The interior regions of Antarctica and Greenland are aseismic: no earthquake larger than body-wave magnitude 4.5-5.0 is known for either, except along coastal zones or continental shelves. Here I advance an explanation for this lack of seismic activity in terms of pressure effects produced by the continental ice sheets that mantle both continents.

Before the International Geophysical Year (IGY) of 1957 there was very little information about the seismicity of either Greenland or Antarctica. The apparent aseismicity could have been due to a lack of detecting seismographs or human inhabitants rather than a true lack of significant earthquakes. With the installation of IGY instruments and of the World-Wide Standard Seismographic Network (WWSSN) in the early 1960s, a global location and detection threshold of body-wave magnitude $m_b = 4.5-5.0$ was achieved¹. In Greenland the WWSSN stations NOR (Nord, 1963-72), GDH (Godhavn, 1962 to present), KTG (Kap Tobin, 1963-80), and DAG (Danarkshavn, 1972 to present), as well as nearby stations in Canada and Iceland, ensure a detection threshold no higher than this, and probably lower, for the entire continent. The WWSSN stations SBA (Scott Base, 1963 to present), SNA (Sanae, 1962 to present) and SPA (South Pole, 1963 to present) monitor the Antarctic continent.

Figure 1 shows the seismicity of Greenland and Antarctica, with elevation contours of the ice sheets. Data for Greenland are taken from Gregerson², Fujita *et al.*³, the NOAA Earthquake Data Base⁴, and the International Seismological Centre⁵. Data for Antarctica are available from refs 4 and 5 only. While the coasts and continental slopes of Greenland exhibit a moderate seismicity rate, even these zones are virtually aseismic in Antarctica: only an $m_b = 4.9$ and an $m_b = 5.1$ event (plus aftershock) have been recorded in more than 20 years of monitoring.

Worldwide seismic activity in the stable continental interiors

of lithospheric plates is quite low compared with linear plate boundaries where oceanic plates are involved, or with distributed or diffuse boundary zones where continents interact. However, a recent systematic study⁶ of these stable continental interiors shows a low but persistent level of seismic activity exceeding $m_b = 5.0$. In rare instances (that is, North America, India, south-east China and Australia), magnitude 7 or greater events in such regions have occurred. The continental crust of Greenland and Antarctica is similar to these other continents in both estimated thickness⁷ and geological age and composition^{8,9}. Their lack of earthquakes cannot therefore be attributed to differences in crustal characteristics and the cause must be sought elsewhere. I propose that the explanation lies with the effect of the thick continental ice sheets on the stress regime of the underlying upper crust. Such an explanation has been both promoted¹⁰⁻¹² and discounted¹³ by others, but without a detailed examination of the mechanism advanced here.

There are at least three ways in which the continental ice sheets of Greenland and Antarctica could affect seismicity: (1) induced elastic-stress changes from the overburden; (2) induced pore-pressure changes; (3) induced thermal stresses. Significant thermally induced stress change at hypocentre depth may be dismissed. The base of both the Greenland¹⁴ and Antarctica¹⁵ sheets is at or near the pressure melt point: water was encountered at the base of the 1969 Antarctic drill hole at 2,164 m. Thus ice sheets insulate the underlying crust from large temperature excursions and maintain the surface temperature near 0 °C. I shall not consider thermal effects further.

The 2,800,000 km³ of Greenland's ice cap covers ~80% of the land area; for Antarctica the respective figures are 12,500,000 km³ and 90%. Ice-raftered debris recovered from very recent deep-sea drilling results in the Weddell Sea¹⁶ indicate that the East Antarctic ice sheet formed between 37 and 35 Myr ago and has existed continuously since at least 13 Myr ago. The West Antarctic sheet apparently formed separately but has been a permanent fixture since ~4.5 Myr ago. Estimates of the time required to form the Antarctic sheet range from 15,000 to 100,000 yr^{17,18}. Hence an ice sheet appears quickly with respect to its duration. It imposes a static load on the underlying crust that would require 600-60,000 Myr to accumulate at the normal tectonic strain rates of $10^{-10}-10^{-12}$ yr⁻¹ that apply to the stable continental interiors of plates¹⁹.

Stress changes due to the weight of an ice sheet and its possible alteration of the crustal pore-pressure regime can be discussed in terms of the Navier-Coulomb failure criterion. Brittle shear failure in rock (an earthquake) is described by $\tau = \tau_0 + \mu\sigma_n$, where τ is the shear stress necessary to induce failure on a fault plane, τ_0 is the inherent shear strength of the fault (the cohesion), μ is the coefficient of friction of the fault surface and σ_n is the normal stress on the fault. Many faults exist in the crust on which failure may or may not be possible for a given stress regime. In the stress diagram of Fig. 2, the locus of shear (τ) and normal (σ_n) stress components on a suite of faults of various orientations is a circle (the Mohr circle) whose centre is the average between the maximum (σ_1) and minimum (σ_3) principal stresses and whose radius is $(\sigma_1 - \sigma_3)/2$. Larger differences between the maximum and minimum principal stresses result in a greater radius of the Mohr circle, meaning that available shear stresses more closely approach the failure stress criterion.

A large body of work studying seismicity induced by the impoundment of reservoirs by dams has been summarized by Simpson²⁰. Reservoirs trigger earthquakes by (1) increasing elastic stresses by incrementing the vertical overburden stress and (2) increasing the hydrostatic pore fluid pressure P such that σ_e , the effective normal stress on a fault, is $\sigma_e = \sigma_n - P$. This latter effect is equivalent to a translation of the Mohr circle along the σ_n -axis to lower values and closer to the failure envelope. The large number of documented cases of reservoir-induced seismicity indicate that continental crust in plate interiors is commonly under high deviatoric stress: it is close

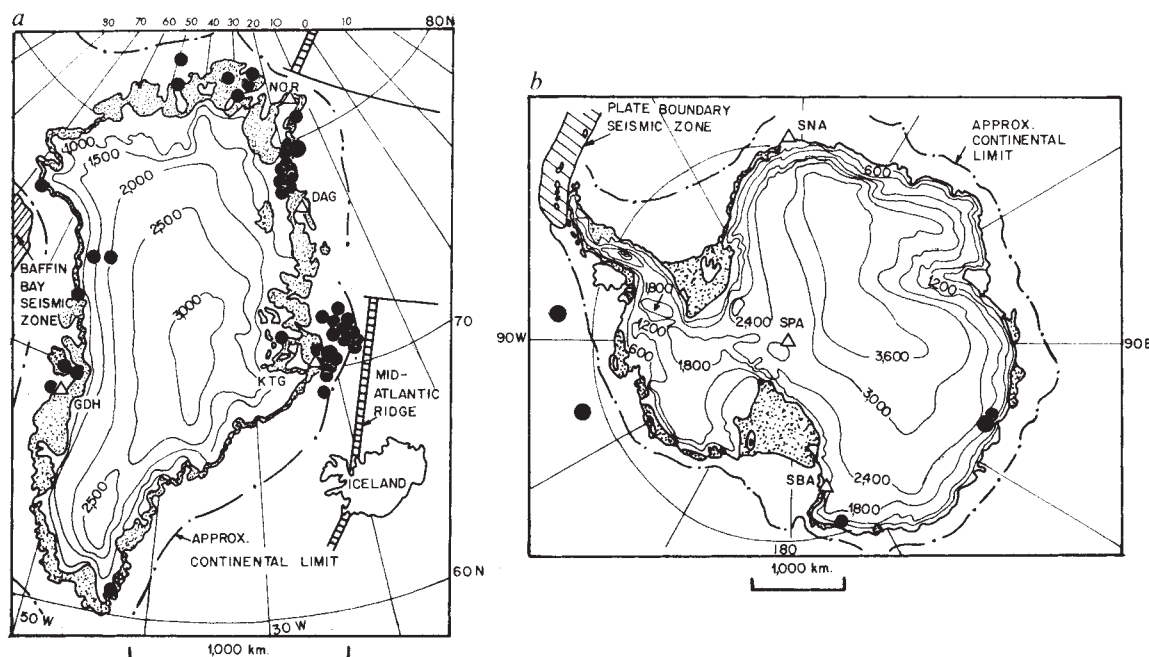


Fig. 1 The seismicity of (a) Greenland and (b) Antarctica. Small dots, magnitude 4.0–4.9; large dots, magnitude ≥ 5.0 . Data are complete to $m_b = 4.5$ for 1963–86, although, if detected, pre-1963 events are shown. Ice-sheet elevation contours in metres and important seismographic stations (triangles) are also shown. Stippling denotes ice-free land.

enough to failure for stress increments on the order of tens of bars or less to induce failure.

For the static load of a reservoir (or an ice sheet) to promote shear failure on faults, the maximum principal stress (σ_1) needs to be vertical, that is, a normal or gravity dip-slip faulting environment exists. The increased static load will increase σ_1 and thereby produce larger shear stresses on favourably oriented faults. (This is equivalent to increasing the Mohr circle radius.) Conversely, in thrust faulting environments where σ_1 is horizontal and σ_3 is vertical, an increase in the vertical stress inhibits failure. I argue that this is so for Greenland and Antarctica, even though no stress data in the form of *in situ* stress measurements or focal mechanisms are available for their interior regions.

The evidence for the existence of a horizontal compressive stress regime in Greenland and Antarctica, although indirect, is fairly compelling. Compilations^{6,21,22} of *in situ* stress measurements, focal mechanisms and stress-sensitive geological features show that most other stable continental interiors are under horizontal compression, yielding thrust-faulting earthquake mechanisms (or strike-slip where σ_1 is still horizontal but σ_2 , the intermediate principal stress, is vertical). An exception is southern and western Africa, where the influence of the East African rift extensional stress regime apparently extends far beyond the mapped rift boundaries. Moreover, global finite-element and finite-difference modelling studies^{21,23} predict horizontal compression for Greenland and for the oceanic portion of the Antarctic plate. Two Antarctic oceanic intraplate focal mechanisms have been obtained^{24,25}, both of which are predominantly thrust faulting solutions.

Although the *in situ* evidence is lacking, the several lines of circumstantial evidence cited above indicate that both the Greenland and Antarctic continental interiors are probably subject to a compressive horizontal stress regime. The imposed weight of the massive ice sheets would therefore suppress earthquake faulting rather than promote it, as is often the case with reservoir impoundment.

A second mechanism whereby ice sheets may reduce seismicity is through control of the crustal pore pressure regime.

The ice would insulate the underlying crust from diffusion of any significant amount of meteoric water from the surface down to the seismogenic portions of the crust. The amount of liquid water that may be present at the ice-sheet base is a thin film^{14,15} and surely insignificant compared with pore-fluid recharge in fluvial environments. It has not been shown that surface fluids or surface-generated transient pressure fronts can reach hypocentre depths, but accumulating data from deep continental drilling²⁶ and conceptual studies²⁷ favour the possibility. If surficial fluid processes are an important control in the pore-pressure regime of the upper crust and therefore important in fault failure, continental ice sheets should significantly inhibit this seismogenic mechanism.

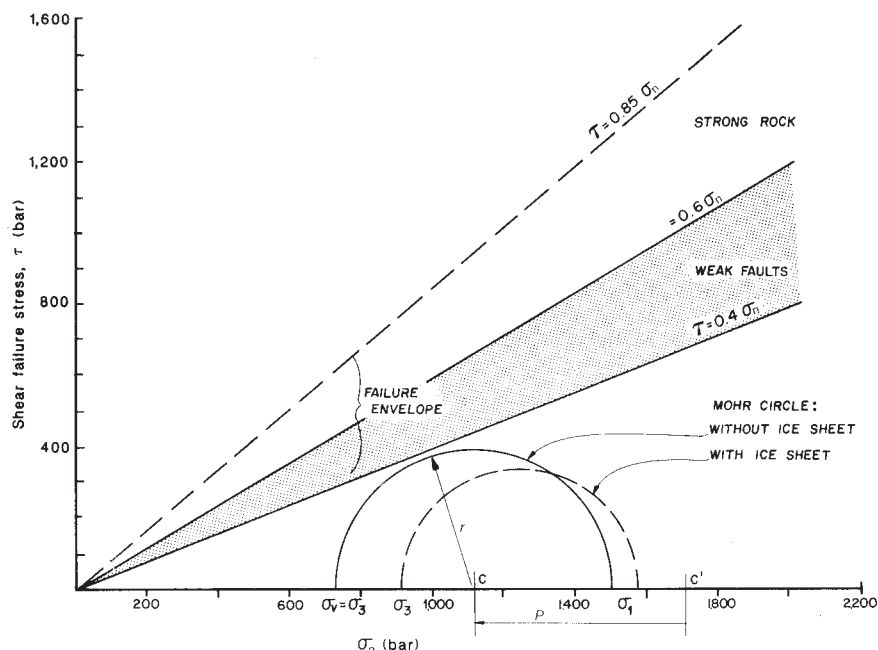
Figure 2 is a hypothetical sub-ice sheet crustal model summarizing the stress-change arguments advanced here. Most parameters and boundary conditions must be assumed or inferred, but the exercise is useful for illustrating the concept of the seismicity suppression mechanism.

A worst-case crustal stress regime is assumed, one in which favourably orientated faults are near the threshold of failure. This assumption is not required, but because other ice-free continental land masses experience significant earthquakes they must contain faults that approach this threshold. The model thus assumes that Antarctica and Greenland would be similar to other continents in the absence of ice sheets.

To approach failure threshold easily, extremely weak faults are modelled with $\tau_0 = 0$, and $\mu = 0.4$ – 0.6 . Laboratory studies²⁸ indicate that for intact rock samples $\tau_0 \approx 0.5$ kbar and $\mu \approx 0.6$ – 0.85 . However, when pre-existing faults with fault gouge are present, μ may range as low as 0.2–0.4 and τ_0 approaches zero^{29,30}. These are the types of faults that easily produce earthquakes.

In accord with the proposed mechanism to suppress earthquakes, these faults are assumed to exist in a thrust-faulting environment (σ_3 vertical). A nominal hypocentre depth of 5 km yields a vertical overburden stress (σ_3) of 1,320 bar (1 bar = 0.1 MPa (mega pascals) in SI units) for a crustal density of 2.7 g cm^{-3} . From some North American drill hole data³¹, a stress ratio σ_3/σ_1 of 0.63 is applied, yielding $\sigma_1 = 2,100$ bar. Pore fluids

Fig. 2 The Navier-Coulomb failure representation of continental crust close to shear failure on favourably oriented faults, for the following conditions: 5 km depth, $P = 590$ bar, σ_3 vertical, $\tau_0 = 0$, $r = (\sigma_1 - \sigma_3)/2$. The effect of a large ice sheet is to move the population of such faults (dashed Mohr circle) away from the envelope of failure criteria. The effect of pore pressure, P , is to move faults closer to failure (C' to C) by reducing σ_n . Continental ice sheets would be expected to inhibit increases in P that result from surface fluid recharge.



at depth, either recharge or connate, are expected to be brines of high specific gravity³¹; thus a hydrostatic pore fluid of high density, 1.2 g cm^{-3} , is included in the model. It will reduce all principal stresses by ~ 590 bar. (A pore fluid at pressure appears to be almost always necessary to bring faults near to failure.)

Upon this crust, modelled so as to require faults on the verge of failure, the static load of an ice sheet is imposed. A density of 0.9 g cm^{-3} is assigned to the ice¹⁴ and a mean thickness of 2 km is taken from Fig. 1. This will increase the vertical σ_3 by ~ 180 bar. To compute the change of the horizontal σ_1 principal stress component, a typical crustal Poisson ratio γ of 0.25 is assigned and the 'lateral constraint' boundary condition, which allows no horizontal deformation, is imposed. This results in $\Delta\sigma_1 = [\gamma/(1-\gamma)]\Delta\sigma_3 \approx 60$ bar. The applicability of this particular boundary condition to the Earth's crust is disputed³². It seems to be reasonable for an external load applied 'instantaneously' ($\sim 10^4$ yr) to a rigid crust many millions of years older. With lateral constraint and hydrostatic pore pressure, $\sigma_1 = 1,570$ bar and $\sigma_3 = 910$ bar. The stress difference between σ_1 and σ_3 that promotes failure is reduced from 770 bar to 660 bar by the ice-sheet load (Fig. 2). Even if an isotropic stress state is assumed ($\Delta\sigma_1 = \Delta\sigma_3$) so that $\sigma_1 - \sigma_3$ remains 770 bar, the static load translates the Mohr circle to higher values of σ_n , that is, towards stability, even though its radius r is not reduced.

In this conceptual model the net effect of a continental ice sheet is to remove faults from the threshold of failure (within 5–10 bar of the Navier-Coulomb failure criterion) and to suppress earthquake faulting (~ 120 bar from failure criterion (lateral constraint); ~ 60 bar from failure (isotropic)). Moreover, restoration to a near-failure condition by increase of pore pressure is inhibited because ice sheets insulate the upper crust from significant surface-fluid recharge. An increase in pore pressure could occur but would require a substantial reduction of available pore volume and would probably be transient compared with the lifetime of an ice sheet.

Of the two proposed mechanisms for the suppression of seismicity beneath ice sheets, the effect of the static overburden is dominant; it could explain the aseismicity without invoking the pore pressure argument. The proposed mechanism would be more firmly based if σ_3 were shown to be the vertical principal stress in Greenland and Antarctica. If surficial control of the pore pressure regime in the upper crust as proposed by Costain

et al.²⁷ were shown to be a viable mechanism, this too would support the role of ice sheets in controlling seismicity. It would also help to explain the extremely low seismicity of some of the very arid regions of the Earth such as the African Sahara and the Arabian Peninsula.

I thank R. Richardson and A. McGarr for thorough critical reviews, K. Fujita for supplying preprints, and the Electric Power Research Institute of Palo Alto, California, for research support.

Received 15 May; accepted 25 August 1987.

1. Habermann, R. E. *Bull. seism. Soc. Am.* **72**, 93–111 (1982).
2. Gregersen, S. *Bull. geol. Soc. Denmark* **31**, 11–27 (1982).
3. Fujita, K., Cook, D. B., Hasegawa, H., Forsyth, D. & Wetmiller, R. in *The Arctic Ocean of North America* (eds Grantz, A., Johnson, L. & Sweeney, J. F.) Vol. I (Geological Society of America, Boulder, Colorado, in the press).
4. *Earthquake Data Base* (Natl. Geophys. Data Center of the National Oceanic and Atmospheric Administration, Boulder, Colorado, 1985).
5. *Catalogue of Earthquakes* (International Seismological Centre, Edinburgh, 1918–85).
6. Coppersmith, K. J., Johnston, A. C., Metzger, A. M. & Arabasz, W. J. *Working Report for EPRI Project 2556-12* (Electric Power Research Institute, Palo Alto, 1987).
7. Gregersen, S. *Geophys. J. R. astr. Soc.* **22**, 29–39 (1970).
8. King, P. B. *Tectonic Map of North America 1:5,000,000* (G67154) (US Geological Survey, Washington, DC, 1969).
9. Grikurov, G. E., Znachko-Yavorskii, G. A., Kameney, E. N. & Kurinin, R. G. (compilers) *Tectonic Map of Antarctica 1:10,000,000* (Research Institute of the Geology of the Arctic, Moscow, 1978).
10. Voronov, P. S. & Klushin, I. G. *Problemi Arktiki i Antarktika, Sb Statei, Arkticki Antarkt. Nauchn. Issled. Inst.* **12**, 5–13 (1963).
11. Evison, F. E. *N.Z. J. Geol. Geophys.* **10**, 479–483 (1967).
12. Burke, K. et al. *Rifts and Sutures of the World*, (NASA-CR-175201) (State University of New York, Albany, 1978).
13. Sykes, L. R. *Rev. Geophys. Space Phys.* **16**, 621–688 (1978).
14. Patterson, W. S. B. *The Physics of Glaciers* (Pergamon, Oxford, 1969).
15. Flint, R. F. *Glacial and Quaternary Geology* (Wiley, New York, 1971).
16. Kerr, R. A. *Science* **236**, 912–913 (1987).
17. Hollin, J. T. J. *Glaciol.* **4**, 173–195 (1962).
18. Weertman, J. *J. geophys. Res.* **73**, 2691–3010 (1968).
19. Anderson, J. G. *Bull. seism. Soc. Am.* **76**, 273–290 (1986).
20. Simpson, D. W. A. *Rev. Earth planet. Sci.* **14**, 21–42 (1986).
21. Richardson, R. M., Solomon, S. C. & Sleep, N. H. *Rev. Geophys. Space Phys.* **17**, 981–1019 (1979).
22. Zoback, M. L. & Zoback, M. J. *geophys. Res.* **85**, 6113–6156 (1980).
23. Richardson, R. M., Solomon, S. C. & Sleep, N. H. *J. geophys. Res.* **81**, 1847–1856 (1976).
24. Forsyth, D. *Nature* **243**, 78–79 (1973).
25. Okal, E. A. *Earth planet. Sci. Lett.* **46**, 306–310 (1980).
26. Kozlovsky, Y. A. (ed.) *The Superdeep Well of the Koln Peninsula* (Springer, Berlin, 1987).
27. Costain, J. K., Bollinger, G. A. & Speer, J. A. *Seism. Res. Lett.* **58**, 41–64 (1987).
28. Byerlee, J. W. *Pure appl. geophys.* **116**, 615–626 (1978).
29. Wang, C. Y. & Mao, N. *Geophys. Res. Lett.* **6**, 825–828 (1979).
30. Talwani, P. & Acree, S. *Pure appl. Geophys.* **122**, 947–965 (1984/85).
31. Wesson, R. L. & Nicholson, C. *Open-File Report 86-331* (US Geological Survey, Reston, Virginia, 1986).
32. McGarr, A. *J. geophys. Res.* **92**, 4959–4964 (1987).

Poás volcano crater lake acts as a condenser for acid metal-rich brine

Susan L. Brantley*, Andrea Borgia†‡, Gary Rowe*, Jose F. Fernandez§ & Jennifer R. Reynolds‡||

* Department of Geosciences, Penn State University, University Park, Pennsylvania 16802, USA

† Department of Geological and Geophysical Sciences, Princeton University, Princeton, New Jersey 08544, USA

§ Instituto Costarricense de Electricidad, San Jose, Costa Rica

|| Department of Earth Sciences, Dartmouth College, Hanover, New Hampshire 03755, USA

Models of formation of volcanic-hosted metal deposits hypothesize that subsurface brines transport metals to sites of deposition. Laguna Caliente, in the active crater of Poás volcano, central Costa Rica, consists of acid-sulphate-chloride brine with extreme pH (0.0). Lake water sampled between 9 and 12 January 1987 at eight sites within the 40-m deep, 210-m wide lake revealed temperatures between 58 and 64 °C, and densities (measured in the laboratory at 60 °C) of $1.0575 \pm 0.0015 \text{ g cm}^{-3}$. The water's extreme acidity reflects the dominating input of acid fumaroles. The presence of precipitated silica, gypsum and sulphur as sediments reflects lake-water equilibrium with those phases. Mass balance considerations require that dense brines percolate downwards from the lake into the volcano. These may act as a source of metal-rich brine feeding an underground ore-forming hydrothermal system.

Of the approximately 40 active volcanoes in Central America, roughly one fifth host acidic crater lakes (M. Carr, personal communication). Laguna Caliente, situated in a small pit crater within the main crater of Poás Volcano, Costa Rica, is the most acid and sulphur-rich lake reported¹. The most recent lava eruption at Poás occurred in 1953, at which time the crater lake totally disappeared, not to reappear until after 1965. Phreatic explosions, consisting of frequent irregularly timed geyser-like mud plumes erupting from the lake, resumed in June of this year after eight years of inactivity².

To investigate the chemical homogeneity of the lake, we designed a sampling raft to collect water and measure temperature at depth (Fig. 2a). Our water samples complement earlier near-shore samples which we also summarize³ (Table 1). Chemical compositions for near-shore samples from March 1985 and filtered mid-lake samples from January 1987 showed similar chemical compositions (see examples in Table 1). Unfiltered lake water showed consistent density at all depths at mid-lake, and slightly lower densities along the eastern shoreline (Fig. 2b). Measurements of temperature variation with depth showed 1.2 °C warmer temperatures at mid-depths in the centre of the lake (Fig. 2c). Sediments (predominantly chemical precipitates of amorphous silica, elemental sulphur and gypsum) at the eastern shoreline measured 6 °C cooler than overlying water.

Our temperature data show the same patterns as reported for the lake in 1984 (X. Neshyba *et al.*, in preparation). At that time, bottom sediments at the centre of the lake were 1.6 °C hotter than bottom waters. These data indicate that the lake is heated predominantly by fumaroles passing through bottom sediments and through the rocks of the dome on the southern side. The observed discoloration at lake centre and the site of observed phreatic activity in the southern section (Fig. 1) may be major sites of convective upwelling. Convection explains the density, temperature and chemical homogeneity. Lower density values and sediment temperature along the eastern shore may

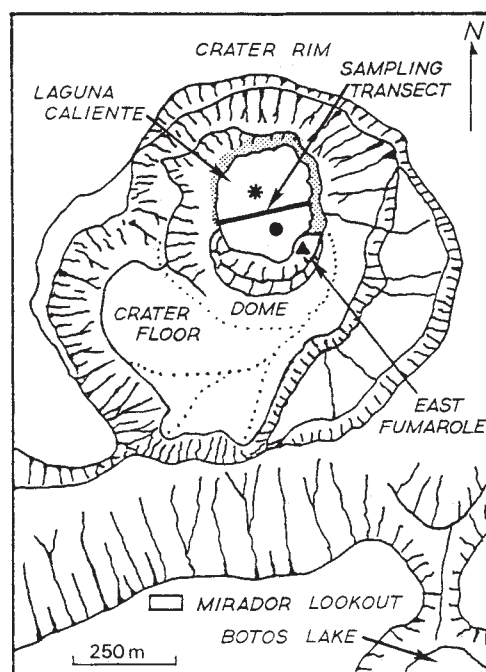


Fig. 1 Map of main crater of Poás Volcano, indicating location of Laguna Caliente, the eastern fumarole, and location of our sampling transect. A site of occasional lake-water discoloration (*) and the site of phreatic explosions (●) are also noted. Dotted lines represent the bed of the intracrater ephemeral river. Sketched lake area is approximately as observed in January 1987, when the level was approximately 7 m lower than January 1985. Between 7 and 10 January 1987, no rain fell and we observed that the lake-water depth fell at the shoreline by about 12 cm d⁻¹. During this period, the lake level had reached the lowest level of the past ten years (J. Barquero, personal communication), revealing extensive layered lake sediments (stippled area) and occasional hollow yellow tubes, approximately 3 cm in diameter and 20 cm in length, composed of radiating, orthorhombic sulphur crystals.

indicate input from groundwater associated with the ephemeral stream (Fig. 1).

Charge balance in the brine can be approximated to first order as $[\text{H}_3\text{O}^+] = [\text{Cl}^-] + [\text{HSO}_4^-]$, where the most highly concentrated species in the lake is the hydronium ion. The extremely low pH of the water results from the continual condensation of fumarolic acid vapors; acidity is enhanced by continual evaporation of water, producing isotopically heavy acid-sulphate-chloride brine (Fig. 3). Because lake water has maintained consistently low pH since at least 1979, pH-buffering crater rock dissolution reactions must proceed at slow rates compared with the rate of fumarole input. Banding in precipitated lake sediments reflects fluctuations in saturation index of gypsum and silica related to short- or long-term variations in temperature, rainfall or volcanic activity.

To quantify lake inputs and outputs (Fig. 3), we calculate a water mass, heat and chloride mass balance between 1985 and 1987. The net rate of water loss, R_t , is determined by the rates of water loss through evaporation (R_e) and subsurface outflow (R_{of}), and rates of water addition from rainfall (R_r) and fumaroles (R_f):

$$R_t = -R_e - R_{of} + R_r + R_f \quad (1)$$

If conduction of heat between lake and wall rocks, loss of heat from the lake through air convection and radiation of heat from the lake are minimal (as found for the crater lake of Soufriere)⁴, the rate of heat change, Q , will also be determined by the rates of heat loss through evaporation, outflow and rainfall, and the

‡ Present addresses: Smithsonian Institution, Department of Mineral Science, NMNH, Washington DC, 20560, USA (A.B.) and Lamont-Doherty Geological Observatory, Palisades, New York 10964, USA (J.R.R.).

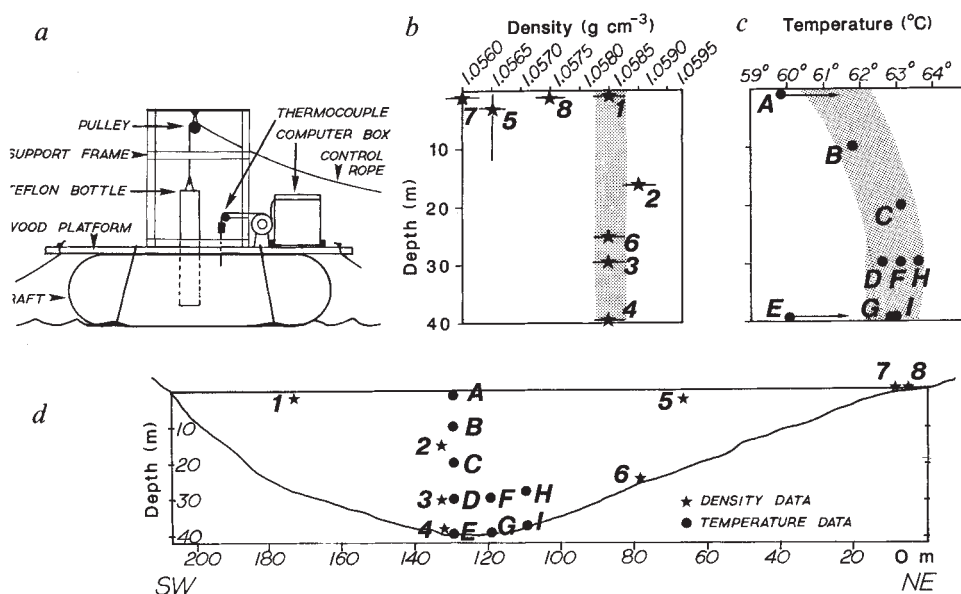


Fig. 2 *a*, The sampler consisted of a small rubber raft with a wood and metal frame rigged with pulleys. The raft was manipulated from opposite shores, using two polypropylene ropes strung across the lake. A third line was used to lower a weighted Teflon bottle to the desired depth where the bottle was triggered to fill. Lake depth was determined at points noted by measuring when the sampler weight hit bottom. In independent measurements, temperature was recorded with a Type-K thermocouple (accurate to $\pm 0.25^\circ\text{C}$) suspended from the raft and attached to an onboard computer. All undiluted samples formed a silica precipitate upon cooling, so samples were diluted in the field for subsequent chemical analysis. *b*, Density data for lake samples. All densities were measured on unfiltered samples in the laboratory at 60°C . Stippled area shows density profile at mid-lake. *c*, Temperature data for mid-lake as measured 9 January 1987. Where no definite temperature is noted (sites A and E), the thermocouple never fully equilibrated. Stippled area shows estimated temperature profile at mid-lake. *d*, Locations of temperature measurements and sample locations along the 210-m transect.

Table 1 Chemical composition of lake system components

	Lake water* (p.p.m.)	Lake water† (p.p.m.)	Lake water‡ (p.p.m.)	Fumarole condensate§ (p.p.m.)	Rio Agrio (p.p.m.)	Rain water¶ (p.p.m.)	Crater lavas# (wt. %)	Lake sediments** (wt. %)
Date	24 Jan. 85	8 Jan. 87	8 Jan. 87	1 Mar. 85	14 Feb. 87	—	—	—
<i>T</i>	$\sim 44^\circ\text{C}$	$\sim 60^\circ\text{C}$	$\sim 60^\circ\text{C}$	584°C	NA	NA	NA	NA
pH	0.0	0.0	0.0	NA	2.2	4.3	NA	NA
F	1,390	1,460	1,420	410	24.1	0.5	NA	NA
Cl	25,400	30,000	31,300	13,600	670	7.1	NA	0.04
SO ₄	49,500	64,400	64,800	3,420	1,660	11.4	NA	16.5 (Σ S)
Na	770	504	500	7	66.4	NA	2.15	0.05
K	300	236	234	5	24.0	NA	1.25	0.06
Ca	1,200	2,860	2,400	31	125	NA	5.50	14.3
Mg	790	529	536	2	82.1	NA	2.39	0.1
Al	3,000	2,000	2,040	19	218	NA	8.95	0.7
Si	50	164	190	60	47	NA	25.0	16.4
Fe	1,400	978	971	10	64.7	NA	6.28	0.16
Unfiltered density (g cm^{-3} at 60°C)	NA	1.0590	1.0585	NA	1.0025 (at 25°C)	NA	NA	NA
δO^{18}	+5.5	12.3	NA	+3.0	NA	+2 to -4 (est.)††	NA	NA

NA, not analysed.

* SE lake shore, surface³.

† Mid-lake, 15 m depth (sample 2 on Fig. 2*d*).

‡ Mid-lake, 40 m depth (sample 4 on Fig. 2*d*).

§ Sampled from eastern fumarole³, 1 Mar. 1985.

|| Sampled 14 Feb. 87, NW flank of volcano, flow $> 100 \text{ l s}^{-1}$.

¶ Sampled at Mirador, average value for 17 samples Nov. 1984–Mar. 1985¹¹.

Average of 14 summit lava flows¹².

** Sampled from SE shore, 1985.

†† Values for rainfall for similar latitude (Barbados)¹³.

rate of heat input from fumaroles:

$$Q = -\Delta H_e R_e - h_{of} R_{of} - \Delta H_f R_f + \Delta H_l R_l \quad (2)$$

where ΔH_e is heat of vaporization, h_{of} is specific enthalpy of lake water, ΔH_f is heat loss in warming rainwater to lake temperature and ΔH_l is the heat of condensation of fumarolic water vapour to liquid at lake temperature. This calculation assumes that all transfer of heat from fumaroles to the lake occurs as

condensation; if heat is also transferred conductively, this assumption will tend to overestimate water input from fumaroles. From January 1985 to January 1987, the lake decreased in depth from approximately 47 m to 40 m, and increased in average temperature from 44°C to 60°C , indicating that $Q = -2 \times 10^6 \text{ W}$ and $R_l = -4 \text{ kg s}^{-1}$. This calculation is based on a lake volume modelled as a cone of depth 40 m and diameter 210 m capped by a cylinder of variable height.

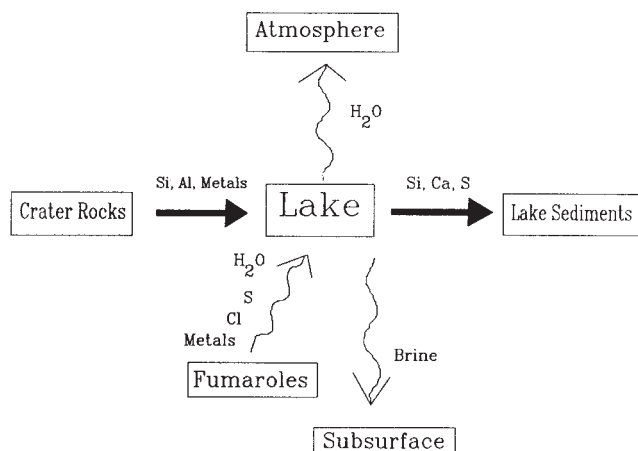


Fig. 3 Simplified box model of Laguna Caliente inputs and outputs. The lake acts as a condenser system, in which H_2O - S - Cl fumarolic fluid enters, condenses and then concentrates through evaporation of water to the atmosphere (vertical arrows show net direction of fluid fluxes). Crater rocks erode into the lake by rain-fed rivers, dissolve and cause supersaturation with respect to gypsum and silica (horizontal arrows). Net result is accumulation of rock- and fumarole-derived metals and anions in lake-water brine, which we hypothesize then seeps downwards into the volcano. Chemical analyses of lake water reveal concentrations between 100 and 3,000 p.p.b. of Cr, Co, Ni, Cu, Zn, Cd and Pb, and concentrations between 10 and 2,000 p.p.m. of Ti, V, Mn and Fe. Note that oxygen isotope values of lakewaters are significantly heavier than fumaroles and estimated rainfall values, (Table 1), indicating that downward-flowing brines provide a source of isotopically-heavy waters to the subsurface groundwater system.

Data from Table 1 for lake chloride concentrations for samples in 1985 and 1987 show that lake chloride concentration $[\text{Cl}]_l$, remained approximately steady at 0.03 kg (Cl) per kg (H_2O), despite fumarolic input ($[\text{Cl}]_f = 0.01 \text{ kg (Cl) per kg (H}_2\text{O})$) and overall decrease in lake volume. We postulate that brine outflow occurs as seepage through the sediments into the underlying crater rocks as recently suggested for other crater lakes⁵. Mass balance on chloride requires that the net rate of change in total lake chloride ($=[\text{Cl}]_l R_t$) represents a balance between fumarole input and brine outflow as seepage:

$$[\text{Cl}]_l R_t = [\text{Cl}]_f R_f - [\text{Cl}]_l R_{\text{of}} \quad (3)$$

The hypothesis of brine seepage is a necessary consequence of the fact that, despite continual inputs, no outlets from the lake have been observed for dissolved species other than Si, S and Ca (Fig. 3). The steady lake chemistry and low pH observed over the last six years^{1,3} require an outlet for the brines.

Solving equations (1)–(3) using appropriate enthalpy values and $R_r = 60 \text{ kg s}^{-1}$ (based on an average rainfall of 4 m yr^{-1} over a catchment area estimated from aerial photographs to be twelve times the lake area⁶), we calculate $R_e \approx 100 \text{ kg s}^{-1}$, $R_{\text{of}} \approx 40 \text{ kg s}^{-1}$, and $R_f \approx 80 \text{ kg s}^{-1}$. Our estimate of brine outflow from the lake is approximate, since we have overestimated contribution of fumarolic condensation, and neglected both vaporization of HCl at the lake surface and temporal and spatial variations in fumarolic chloride content. However, recasting the rates as lake level drop per day, we can calculate drop during periods of no rain ($R_r \approx 0$, neglecting groundwater input): $\sim 10 \text{ cm per day}$. This calculated rate of drop compares well with the drop of 12 cm per day observed during the rainless period of 9–12 January 1987.

Similarities between Poás lake and the well-described lakes of Ruapehu and Chichón, including similar water and sediment chemistries, complete water mixing and layered sediments, suggest that such crater lakes commonly evolve along the same chemical lines^{7,8}. Evidence from Poás suggests that such lakes act as condensers to produce acid, metalliferous and isotopically

heavy brine which then percolates downward into the volcano (Fig. 3). Downward-percolating brine must either vaporize and re-rise as fumaroles, remain stored as groundwater, or reissue from the flanks of the volcano as a low-pH high-salinity spring (for example, Río Agrio, Table 1). The downward percolation plays an important part in the mass balance and should provide a significant input of metals (for example, tens of grams of Fe per second) into the local groundwater system. Indeed, circulation of hot, dense metal-rich brine is a necessary component of the accepted models for generation of metal deposits within volcanic stockworks⁹. We expect that downward-flowing brines at Poás may be causing subsurface acid-sulphate alteration and possibly contributing to the generation of heavy metal ore deposits¹⁰.

We thank Barry Voight, Jorge Barquero, Paul Briggs, Bob Stallard, Barbara Barreiro and Richard Stoiber. Support was provided by the NSF Latin American Cooperative Program, Andrew W. Mellon Grant, Richter Memorial Trust Research Grant, Penn State, Princeton University, Dartmouth College, Instituto Costarricense de Electricidad and the Sociedades Parques Nacionales.

Received 22 June; accepted 29 September 1987.

1. Casertano, L. *et al. Geof. Int.* **24–2**, 315–332 (1985).
2. Barquero, J. *SEAN Bull.* **12** (8) (in the press).
3. Reynolds, J. *thesis*. Dartmouth Coll. (1985).
4. Shepherd, J. B. & Sigurdsson, H. *Nature* **271**, 344–345 (1978).
5. Henley, R. W. in *Geology and Geochemistry of Epithermal Systems* (eds Berger, B. R. & Bethke, D. M.) 1–24 (Society Economic Geologists, El Paso, Texas, 1985).
6. *Anuario Meteorológico*, 1976–83 (Instituto Meteorológico Nacional, San Jose, Costa Rica).
7. Giggenbach, W. *N.Z. J. Sci.* **17**, 33–45 (1974).
8. Casadevall, T. J. *et al. J. Volcan. geotherm. Res.* **23**, 169–191 (1984).
9. Sillitoe, R. H. *Econ. Geol.* **68**, 799–815 (1973).
10. Hayba, D. O., Bethke, P. M., Heald, P. & Foley, N. K. in *Geology and Geochemistry of Epithermal Deposits* (eds Berger, B. R. & Bethke, P. M.) 129–268 (Society Economic Geologists, El Paso, Texas, 1985).
11. del Rosario Alfaro, M., Fernandez, E., Barquero, J., Rodriguez, J. J. & Rodriguez, M. *Bol. Volc.* **17**, 15–22 (1986).
12. Prosser, J. & Carr, M. J. *Volcan. Geotherm. Res.* (in the press).
13. Dansgaard, W. *Tellus* **16**, 436–468 (1964).

Late-glacial mammoth skeletons from Condover, Shropshire, England

G. R. Coope* & A. M. Lister†

* Department of Geological Sciences, University of Birmingham, Birmingham B15 2TT, UK

† Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK

We report the discovery of abundant skeletal remains of mammoth, *Mammuthus primigenius* (Blumenbach) from late-glacial sediments filling a kettle-hole near Shrewsbury, Shropshire in England. Radiocarbon dated to $\sim 12,800 \text{ yr BP}$, they are by far the latest dated remains of mammoth from Britain, extending its known occurrence there by around 5,000 yr. They provide clear evidence that the full-glacial episode of the last cold stage did not bring about the final extinction of the mammoth in western Europe, but that the species returned in the late glacial before its ultimate demise. The finds represent the most complete mammoths found in Britain, and include the largely complete skeleton of an adult together with partial skeletons of at least three juveniles.

During removal of superficial deposits at the ARC (Western) gravel pit at Norton Farm, Condover (5 km south of Shrewsbury; map ref. SJ498075) in September 1986, numerous mammoth bones were thrown onto spoil heaps. Bone associations suggest that articulated skeletons had occurred *in situ*, but unfortunately their taphonomic relationships were largely destroyed during excavation. Systematic sorting of the spoil heaps by teams of volunteers organized by Shropshire County Museums Service

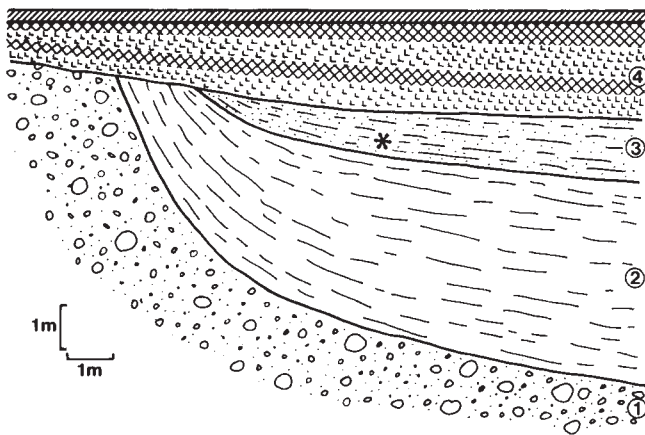


Fig. 1 Diagrammatic section of the Norton Farm kettle-hole, showing the position of the mammoth skeletons (asterisk). (1) Glacial sands and gravels; (2) pink laminated clay; (3) grey sandy clay; (4) Flandrian peats and clays.

resulted in the recovery of all the major postcranial bones of an adult mammoth, and an assortment of juvenile skeletal elements.

The Condoover area is within the ice limit of the late-Devensian ice advance which probably reached its maximum some time between 20,000 and 18,000 yr ago¹. The topography is hummocky with melt water channels and numerous enclosed basins, some with standing water, others now infilled. The hummocks are composed of glacial sands and gravels with interbedded tills, and adjacent to the margins of basins the bedding is steeply inclined toward the centre, presenting evidence of collapse due to the melting of buried ice. The basins are therefore interpreted as kettle-holes.

The kettle-hole at Norton Farm pit was completely filled with sediment (Fig. 1). Four metres of basal pink laminated silty clay graded rapidly into overlying grey sandy organic clays. The latter deposit, in which the mammoth bones occurred, was observed to be at least 1 m thick, and was probably thicker in the centre of the depression, but unfortunately, this had all been removed prior to investigation. Directly above the mammoth horizon were subhorizontal sets of Flandrian (postglacial) peats and clays which sealed the hollow and were built up to the present land surface. The stratigraphic context thus leaves little doubt that the sediments enclosing the mammoth remains were of late-glacial age.

The adult mammoth is represented by the lower jaws and teeth (Fig. 2a), vertebral column, rib cage, and all major limb bones. Portions of tusk and many of the foot bones have also been recovered, but the cranium and tail vertebrae are missing. All of the bones can be re-articulated, and there is no evidence that more than one individual is represented. The size and wear stage of the lower teeth suggests an age of 30–32 yr (Fig. 2, legend). Skeletal height, based on limb bone lengths, is estimated at 310 cm, falling near the upper end of the range of other Eurasian adult *M. primigenius* skeletons (258–320 cm, $n = 8$; ref. 2), and corresponding to the probable male sex of the individual. Allowing for muscle and skin, shoulder height in life would have been ~340 cm.

The existence of three juveniles is demonstrated by three largely complete lower jaws. Two of these are of almost identical size; the third (Fig. 2b) is somewhat larger. Dental wear indicates an age of 3–4 years for the two smaller individuals, and 5–6 years for the larger (Fig. 2 legend). Other juvenile specimens include ~50% of a cranium with teeth, 17 partial vertebrae, 24 ribs, and 17 limb bones. Not all of these can be allocated to particular individuals, but two preserved ilia (Fig. 3) differ in a way which suggests that the two younger individuals were of

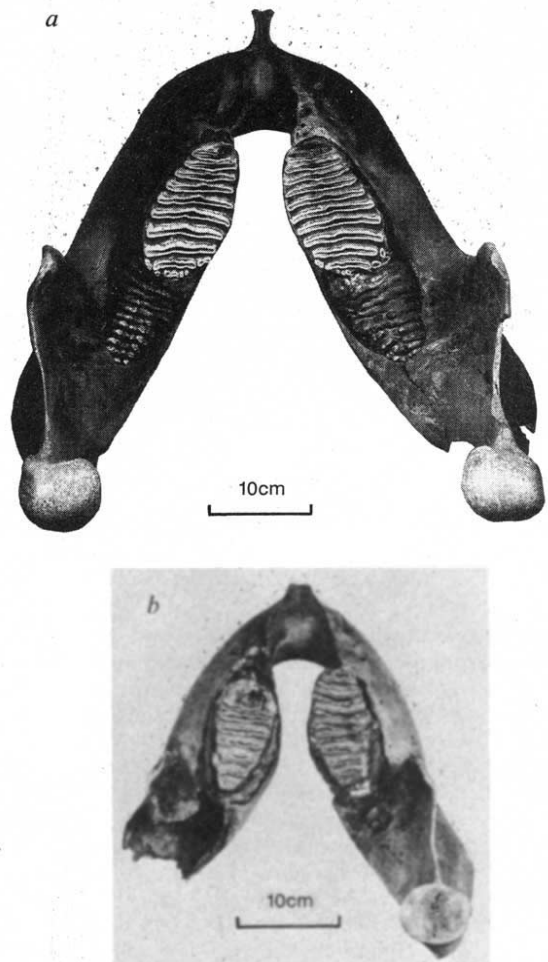


Fig. 2 Lower jaws of mammoths from Condoover, in occlusal (dorsal) view: *a*, adult, *b*, older juvenile. The adult jaw contains two molar teeth in each hemi-mandible, the anterior in advanced wear, the posterior just coming into wear anteriorly. By comparison with extensive published data on European mammoths¹¹, the size of these teeth indicates that they are permanent molars M_2 and M_3 . By analogy with both living African elephant (*Loxodonta africana* (Blum.))¹² and Asian elephant (*Elephas maximus* L.)¹³, this dental stage suggests an age at death of approximately 30–32 yr. In the juvenile jaw (*b*), milk-molar dP_4 is in full wear and M_1 is in place behind it but unworn, suggesting an age of ~5–6 yr^{11–13}. The left M_1 and articular condyle of the juvenile are missing. The two younger juvenile jaws (not illustrated) each contain a worn vestige of dP_3 , and dP_4 in early wear, implying an age of 3–4 yr^{11–13}. These age deductions are supported by the very similar estimates obtained independently from the African and Indian elephant data, and by the similarity in body size of these species to mammoths.

opposite sexes. Detailed descriptions of all the adult and juvenile remains will be given in a full report, to be published elsewhere.

The stratigraphy of the site provides clues to the taphonomy of the skeletons. By the time of deposition of the grey sandy clays, collapse structures round the perimeter of the depression had produced very steep sides, which would have been lined with the pink laminated clay. It is clear from the observed dip of the marginal sediments (Fig. 1) that slopes of at least 40° to the horizontal were present at the time when the mammoths gained access to the hollow, and escape must have been virtually impossible. We envisage the trapped animals either becoming mired in the tenacious deposits, or dying of starvation near the side of the hollow. Rapid burial, or at least submergence, is suggested by the lack of any peri-mortem injury or post-mortem scavenging of the bones.

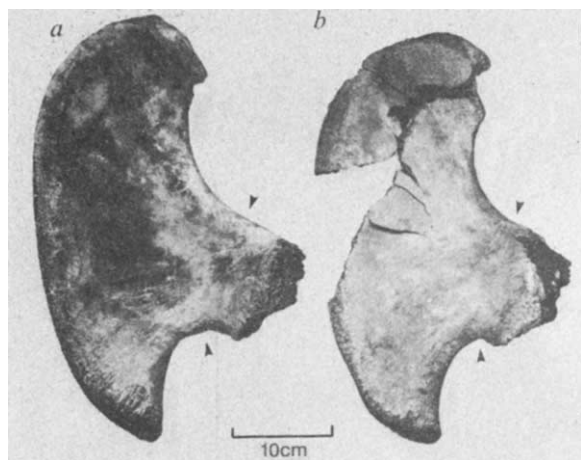


Fig. 3 Left ilia of two juvenile mammoths from Condober, in postero-lateral view. They are of very similar size, but one has a longer shaft (arrowed) and more extended medial and lateral borders (a) than the other (b), suggesting that they are female and male, respectively.

A rich assortment of fossil plants and insects has been recovered from the grey sandy clay (mammoth horizon). These are now being investigated, but their interpretation is complicated by evidence of mixing within the deposit, perhaps due to bioturbation by the trapped mammoths. A precise picture of the contemporary environment may therefore be difficult to reconstruct. The sinuses of the juvenile skull, and two of the juvenile mandibles, however, contained numerous blowfly puparia (*Phormia terraenovae* Robineau-Desroidy), as well as abundant remains of the dung beetle, *Aphodius prodromus* Brahm, both rare in the surrounding matrix. As the present distribution of both species is temperate to boreal, but not arctic, a relatively mild climate is implied.

Radiocarbon dating of collagen³ from the adult mammoth was undertaken both by accelerator mass spectrometry at Oxford University, and by proportional counting of CO₂ at Birmingham University. Using fragments from the same piece of tusk, the two laboratories produced dates of 12,700 ± 160 yr BP (OxA-1021) and 12,920 ± 390 yr BP (Birm-1273), respectively.

Taken together, the radiocarbon dates, stratigraphy, and limited biological evidence suggest that the mammoths lived during an early part of the climatic amelioration which followed the maximum glacial episode, and which preceded the Younger Dryas stadial (~11,000 yr BP) and subsequent postglacial warming (~10,000 yr BP)⁴.

The Condober remains constitute the first clear evidence that mammoths occurred in Britain after the maximum of the last glaciation. They extend the known British occurrence of the species by around 5,000 yr, the previous latest record being from Cae Gwynn Cave, Clwyd, dated to 18,000⁺¹⁴⁰⁰₋₁₂₀₀ yr BP (Birm-146), in deposits overlain by till of the Irish Sea ice sheet⁵. Also, the assemblage of adult and juvenile skeletons is unique in Europe, and the adult skeleton is the best-preserved mammoth of any age yet found in the British Isles. The Aveley mammoth⁶, considerably older, is of comparable completeness, but is in a poorer state of preservation.

Recently, radiocarbon dating of two mammoth finds from continental Europe has pointed to the presence of mammoths there after 13,000 yr BP. These are from Praz Rodet, Switzerland (12,270 ± 210 yr BP: Ly-877) (ref. 7) and Etiolles, France (12,000 ± 220 yr BP: Ly-1351) (ref. 8). Also, palaeolithic engravings and bones of mammoths occur in layers dated to ~12,500 yr BP at Gönnersdorf, W. Germany⁹. These constitute evidence against the widely-held view that mammoths were restricted to Siberia and North America by this late period, having become extinct earlier in Europe¹⁰. The Condober

remains provide important support for the revision of this view. Their excellent state of preservation, clear stratigraphic context, and associated ¹⁴C dating, are a unique combination among the European finds referred to this period. They indicate that we have to look beyond the main Devensian ice advance, to events in the late glacial, for the cause(s) of extinction of the mammoth in Europe.

We are grateful to Mrs Eve Roberts, whose discovery of the bones saved them for scientific study; Mr G. I. McCabe of the Shropshire Museums Service for his assistance and enthusiasm; and ARC (Western) Limited for access to the site and fossils. A.L. acknowledges the support of the Natural Environment Research Council.

Received 30 July; accepted 20 October 1987.

1. Bowen, D. Q. *Geol. J.* **8**, 207-224 (1973).
2. Garutt, W. E. *Das Mammut* (Kosmos, Stuttgart, 1964).
3. Gillespie, R., Hedges, R. E. M. & Wand, J. O. *J. Archaeol. Sci.* **11**, 165-170 (1984).
4. Lowe, J. J., Gray, J. M. & Robinson, J. E. (eds) *Studies in the Late-Glacial of North-west Europe* (Pergamon, Oxford, 1980).
5. Rowlands, B. M. *Nature phys. Sci.* **230**, 9-11 (1971).
6. Stuart, A. J. *Pleistocene Vertebrate in the British Isles* (Longman, London, 1982).
7. Evin, J., Marien, G. & Pachiaudi, Ch. *Radiocarbon* **18**, 60-88 (1976).
8. Evin, J., Marien, G. & Pachiaudi, Ch. *Radiocarbon* **21**, 405-452 (1979).
9. Bosinski, G. *Der Magdalenian-Fundplatz Gönnersdorf*, Vols 1-6. (Steiner, Wiesbaden, 1976-78).
10. Berglund, B. E., Hakansson, S. & Lagerlund, E. *Boreas* **5**, 177-191 (1976).
11. Aguirre, E. E. *Estudios Geol.* **24**, 109-167 (1968).
12. Laws, R. M. E. *Afr. Wildlife J.* **4**, 1-37 (1966).
13. Roth, V. L. & Shoshani, J. *J. Zool.* (in the press).

High population persistence in a system with high turnover

Thomas W. Schoener & David A. Spiller

Department of Zoology, University of California, Davis, California 95616, USA

The concept of turnover—the repeated extinction and immigration of populations in ecological communities—has been prominent for many years¹. Here we show that a system of island orb spiders has consistently high turnover during a five-year period. This might seem to imply that most populations are ephemeral, but ~50 % of the populations present in the system at any one time are persistent over at least the moderately long term. Curves of persistence through time for ensembles of populations are more concave than exponential curves, indicating highest extinction rates in young populations; species vary greatly in the slope of this population 'persistence curve'. In any one-year period, small populations become extinct much more frequently than larger populations. Over longer periods, however, larger populations do not diminish and become extinct; instead, smaller populations re-immigrate and become extinct repeatedly. The total number of individuals belonging to populations becoming extinct during intervals of up to five years is small compared to those persisting. Thus, a layer of small, ephemeral populations is superimposed against a backdrop of large, persistent populations.

Turnover (relative) over a unit time interval, $t_2 - t_1$, is²⁻⁵

$$100 \times \frac{(\text{extinctions of species already present} + \text{immigrations of new species})}{(\text{number of species at } t_1 + \text{number of species at } t_2)}$$

Earlier, we reported that mean species turnover (relative²⁻⁵) in orb spiders on ~100 Bahamian islands⁶ from 1981 to 1982 was high compared to typical values from the literature⁷. Subsequent data confirm the high turnover in this system and show a remarkable constancy in both mean annual turnover (the average over all islands) and its between-island variability: means (standard deviations) are 34.1 (35.4), 29.1 (38.9), 31.3 (33.7), 37.2 (36.8) and 31.0 (34.6) % yr⁻¹ (respective one-year intervals from

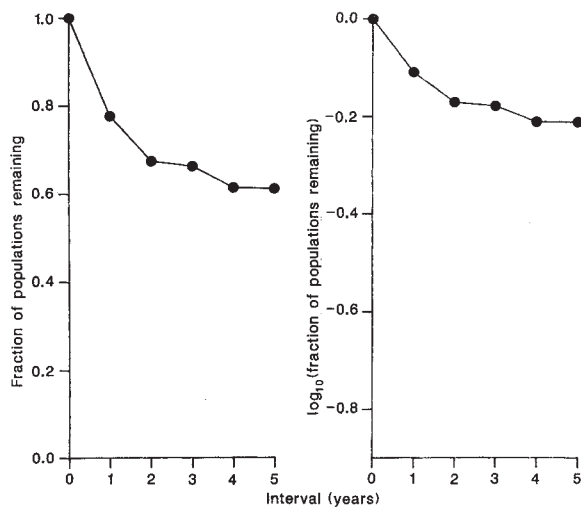


Fig. 1 Population persistence curves for all species of orb spiders combined on 108 islands of the central Bahamas. Curves give the fraction of populations remaining after n years on an arithmetic (left) and semilogarithmic (right) scale. All possible n -yr intervals are used (see text). For successively longer time periods, fewer n -yr intervals are available to compute the curve; the single 5-yr interval from 1981 to 1986 determines the final point. Islands were censused yearly during April–May, 1981–86 (thus our mean annual turnover values may be an underestimate³⁰). Typically, a team of 2–4 researchers inspected all vegetation and other surfaces on an entire island for orbs; a few larger islands were sampled rather than censused completely (see refs 6, 10, 16 for further details). Median vegetated area was $\sim 250 \text{ m}^2$. Islands ($n = 15$) used in introduction experiments in 1983 (ref. 10) are included.

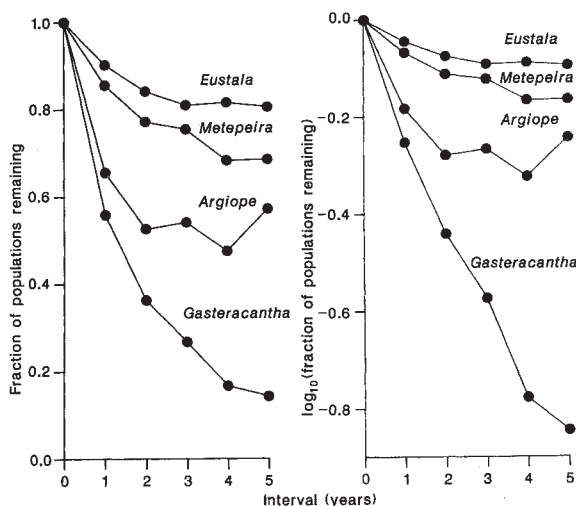


Fig. 2 Population persistence curves for the four commonest orb-spider species on 108 islands of the central Bahamas. Details as in Fig. 1.

1981 to 1986). Such high values suggest that the orb-spider system has a largely fugitive aspect, with most species becoming extinct and reimmigrating repeatedly. On the other hand, the constancy suggests a limit to the number of species that can turn over per unit time, possibly because of a core of permanent populations. More generally, potential contradictions exist between the concept of a system with high turnover and both the population-equilibrium view of communities and the safety from extinction of moderately large populations according to demographic-stochastic theory.

To probe such issues, we counted all populations of diurnal orb spiders on 108 small islands in the central Bahamas (near

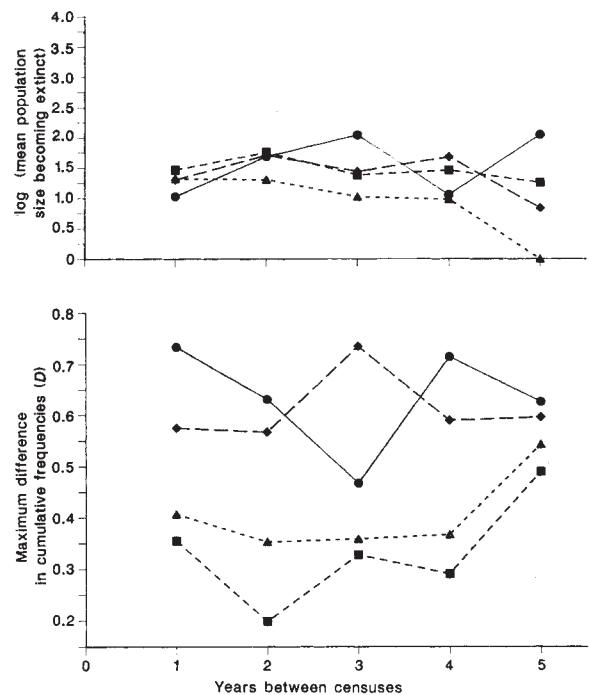


Fig. 3 Top, data bearing on differences between extinct and extant populations as a function of time interval between initial and final censuses. Islands used in introduction experiments in 1983 are included¹⁰. Top, mean size ($\log_2$) of populations found at the beginning but not at the end of the time interval, that is, for extinctions. All n -year intervals combined for each n . $\log_2$ scale was used to conform to Fig. 4 and refs 6, 10. Bottom, maximum difference in cumulative frequency, D , between distributions of extant populations (right histograms, Fig. 4) and those becoming extinct (left histograms, Fig. 4). D is the test statistic for the Kolmogorov-Smirnov 2-sample test¹¹; of the 25 D values illustrated, significance probabilities (one-tailed test) exceed 0.05 for *Gasteracantha* (two- and four-yr intervals), *Argiope* (four-yr intervals) and all five-yr intervals but *Eustala*; of the remaining 17, all statistically significant comparisons, all but *Gasteracantha* (three-yr intervals) and *Eustala* (five-yr intervals) have $P < 0.01$ and nine have $P < 5 \times 10^{-4}$. ●, *Metepeira datona*; ■, *Gasteracantha cancriformis*; ▲, *Argiope argentata*; ◆, *Eustala cazieri* (*Eustala* includes empty webs as well as occupied ones; see refs 6, 10, 16).

Staniel Cay) for six years. Eight species occurred in the system, and of the possible 864 (8×108) species-island combinations, 241 occurred at some time, making this one of the largest data sets available on temporal changes in population numbers^{8,9}.

The most direct way to compare the fugitive and permanent aspects of an ensemble of populations is to construct 'population persistence curves'. These curves give the fraction of those populations observed at a particular time that were still extant after successively longer time intervals, and that did not become extinct and reimmigrate in the intervening time. Because it is arbitrary over which one, two or n -year interval persistence is measured, curves are computed using all possible intervals for the appropriate time span. For example, over 1981–86 five one-year intervals exist, and the population persistence curve gives the fraction of all populations existing at the beginning of any one-year interval still remaining after one year.

Figure 1 shows population persistence curves for all species combined; Fig. 2 shows separate curves for the four commonest species. All curves are very concave and, except for one species, appear to level off sharply at a value greater than (sometimes much greater than) zero. Transformation to a semilog scale reveals that these curves are more concave than exponential curves (a linear relation is expected for the exponential); quadratic regression shows that the concavity is significant for each

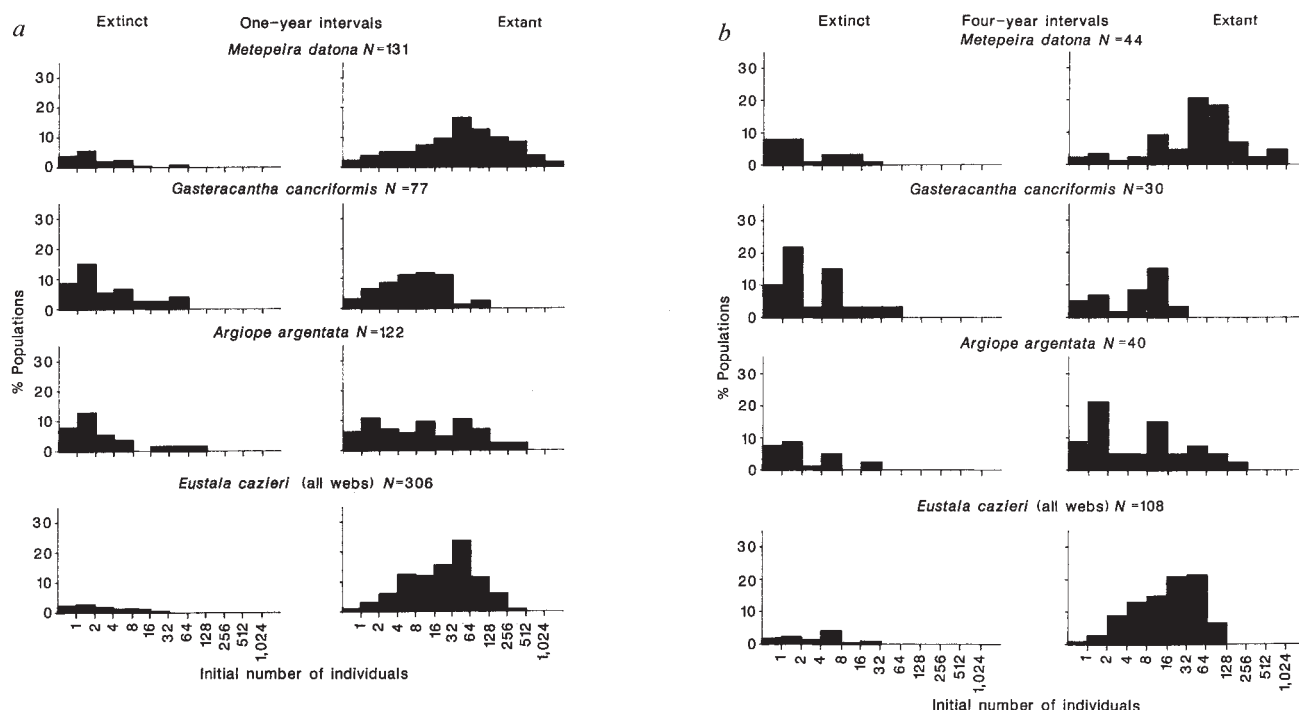


Fig. 4 Histograms of extinct and extant populations over the stated time interval as a function of population size at the beginning of the interval for islands used in Fig. 3. In some cases, populations changed state (became extinct and reimmigrated or vice versa) more than once during the interval—the histograms take into account the initial and final states only. Counts include all individual spiders found in webs. *N*, number of islands examined. Log₂ scale was used to conform to refs 6, 10.

of the five curves (the coefficient of the squared term is positive; one-tailed *t*-test: $P \leq 0.05$). The implication is that the extinction rate for each ensemble of populations is not constant but is higher the shorter the time interval after the initial observation. The curve for all species combined appears to flatten sharply between 0.5 and 0.6, so that ~50% of all species populations occurring in the system at any one time appear to be permanent at least over the moderately long term.

Population persistence curves vary greatly among species, especially in slope and where marked flattening begins, indicating great interspecific variation in the fraction of populations that are permanent. *Eustala* and *Metepeira* show very gentle curves, whereas *Gasteracantha* shows a steeply descending one, while the curve for *Argiope* is somewhat less steep. Looking at the data from a different view, of the total number of islands in which a given species has been found, 17.7% for *Eustala*, 11.5% for *Metepeira*, 36.5% for *Argiope* and 39.4% for *Gasteracantha* had populations which had become extinct and reimmigrated at least once ($N = 79, 42, 52$ and 33, respectively).

Which populations become extinct? Earlier^{6,10} we reported that for the intervals 1981–82 and 1982–83, smaller populations were far more likely to become extinct than larger ones, so much so that the total number of individuals in populations becoming

extinct involved only 3–4% of the total number of individuals in all populations combined. These data suggested that populations turning over are numerically small; however, in the short term populations must pass through small sizes before becoming extinct, and perhaps, over the long term, populations of large initial sizes may diminish and eventually become extinct. The persistence curves presented here suggest the contrary, but we can examine the question directly as follows. Figure 3 shows the initial sizes of populations becoming extinct by the end of increasing time intervals, up to five years. No trend exists for any species. The data on which this graph is based are exemplified in Fig. 4, which gives size distributions of 'extinct' and 'extant' populations over one- and four-year intervals (two-, three- and five-year intervals are similar). For each comparison, populations becoming extinct are much smaller than extant populations. More importantly, no difference between the smaller and larger intervals is apparent: *D*, the maximum difference in cumulative frequencies between 'extinct' and 'extant' distributions (left and right in Fig. 4), is plotted in Fig. 3 (bottom); again, no trend exists. *D* is the test statistic in the Kolmogorov two-sample test¹¹ and most comparisons between 'extinct' and 'extant' distributions are statistically significant (Fig. 3). Finally, except for *Gasteracantha*, the total number of individuals in

Table 1 Total number of individuals in populations becoming extinct as a percentage of *T*, the total number of individuals in all populations combined

	Interval length									
	1 yr		2 yr		3 yr		4 yr		5 yr	
	%	<i>T</i>	%	<i>T</i>	%	<i>T</i>	%	<i>T</i>	%	<i>T</i>
<i>Metepeira datona</i>	0.7	14,350	1.6	11,257	2.3	6,015	1.3	3,217	2.2	902
<i>Gasteracantha cancriformis</i>	28.2	792	36.0	691	38.1	305	54.1	194	61.8	95
<i>Argiope argentata</i>	10.2	3,602	3.7	3,146	5.2	1,041	5.6	643	2.3	88
<i>Eustala cazieri</i>	1.1	11,287	1.7	9,342	1.6	5,468	2.4	2,480	1.1	85
Total all species	2.8	30,226	3.0	24,712	3.1	12,842	3.9	6,545	4.8	1,933

1983 experimental introduction islands (ref. 10) excluded.
Data are combined for each *n*-year interval.

'extinct' populations as a percentage of the total individuals of all populations combined remains very small as the interval of time between initial and final census increases; although the percentage for all spiders combined increases monotonically from one- to five-year intervals (Table 1), its maximum is only 4.8%.

Many data have now been amassed for the Bahamas orb-spider system, nearly all showing that smaller populations are relatively prone to extinction. This tendency is predicted from stochastic-demographic models¹² and is a crucial assumption for certain community-ecological models (for example, the MacArthur-Wilson Equilibrium Model¹), but until now has seldom been demonstrated directly in nature, exceptions being studies of island birds^{3,13}, mainland birds^{3,14}, island lizards^{10,15} and our previous work on island spiders^{6,10}. Our analysis also supports the proposition that populations differ markedly, so that certain kinds of populations are inherently more likely to be smaller than others and hence more extinction-prone; for example, initial sizes of populations becoming extinct are unchanged over intervals of one to five years. Previously we showed, for the orb-spider system, that the populations becoming extinct differentially occurs on small islands⁶, isolated islands⁶ and islands with lizards (unpublished), the latter of which are major predators on and perhaps competitors of spiders^{6,10,16-18}. Our analysis also revealed substantial differences in persistence curves of the four commonest species. Two species that are relatively prone to extinction are *Gasteracantha* and *Argiope*. This is partly because they tend to occur in smaller populations (Fig. 4, Table 1). But even for the same population size, they are more likely to become extinct than other species (Fig. 4), so that characteristics other than population size must be involved. Indeed, both of these species are specialized as adults to occupy relatively large spaces between vegetation, and *Gasteracantha* is specialized for the canopy. Moreover, individuals of one of the more permanent species, *Meteteira*, form colonies with interconnected webs¹⁹, which may enhance individual survival and thereby population persistence.

Despite the high turnover, about half the populations are persistent in the moderately long term in the orb-spider system. Permanent populations may be even more dominant in systems with low species turnover, for instance, those of many vertebrates and higher plants⁷. Consequently, models of how spatial fragmentation affects species diversity, which mostly deal with purely fugitive systems²⁰⁻²⁴, could be expanded to include a permanent as well as a fugitive tier (supporting Hanski's work^{25,26}). Given stochastic-demographic models, which predict that populations attaining a certain large size are effectively safe from extinction under chronic conditions¹², the fact that many permanent populations exist should come as little surprise. What may be surprising, however, is the relative unimportance of populations turning over, as measured by number of individuals and, by implication, biomass or energy (a similar result has been found in continental birds²⁷). This implies that permanent populations must be nearly the sole source of colonizing individuals in the system. Again, this could have implications for fragmentation models.

Apparently unappreciated in MacArthur and Wilson's initial formalization^{1,28} of the turnover concept was the contradiction of large numbers of species becoming extinct with the idea that many populations are in individual equilibrium (also promoted by MacArthur²⁹), as well as the contradiction with stochastic-demographic models. But if in general turnover involves only a subset of fugitive populations, with many others, mostly much larger, being permanent, the contradictions may be reconciled. Thus, although turnover is viewed as somewhat less important, we obtain consistency between certain population- and community-level theory as well as a strong empirical push towards broadening that theory.

We thank two anonymous reviewers for constructive comments. Supported by grants from the NSF and BSR.

Received 14 July; accepted 1 October 1987.

- MacArthur, R. H. & Wilson, E. O. *The Theory of Island Biogeography* (Princeton University Press, Princeton, 1967).
- Diamond, J. M. *Proc. natn. Acad. Sci. U.S.A.* **64**, 57-63 (1969).
- Williamson, M. *Island Populations* (Oxford University Press, 1981).
- Williamson, G. B. *Am. Nat.* **112**, 241-243 (1978).
- Schoener, T. W. *Am. Nat.* (in the press).
- Toft, C. A. & Schoener, T. W. *Oikos* **41**, 411-426 (1983).
- Schoener, T. W. *Oikos* **41**, 372-377 (1983).
- Connell, J. H. & Sousa, W. P. *Am. Nat.* **121**, 789-824 (1983).
- Schoener, T. W. *Am. Nat.* **126**, 633-641 (1985).
- Schoener, T. W. in *Community Ecology* (eds Diamond, J. M. & Case, T.) 556-586 (Harper & Row, New York, 1986).
- Siegel, S. *Nonparametric Statistics* (McGraw-Hill, New York, 1956).
- Goel, N. S. & Richter-Dyn, N. *Stochastic Models in Biology*, Ch. 4 (Academic, New York, 1974).
- Jones, H. L. & Diamond, J. M. *Condor* **78**, 526-549 (1976).
- Beven, G. *Lond. Nat.* **55**, 23-42 (1976).
- Schoener, T. W. & Schoener, A. *Nature* **302**, 332-334 (1983).
- Schoener, T. W. & Toft, C. A. *Science* **219**, 1353-1355 (1983).
- Schoener, T. W. & Spiller, D. A. *Science* **236**, 949-952 (1987).
- Spiller, D. A. & Schoener, T. W. *Ecology* (in the press).
- Schoener, T. W. & Toft, C. A. *Behav. Ecol. Sociobiol.* **12**, 121-128 (1983).
- Levins, R. & Culver, D. C. *Proc. natn. Acad. Sci. U.S.A.* **68**, 1246-1248 (1971).
- Slatkin, M. W. *Ecology* **55**, 128-134 (1974).
- Levin, S. A. *Am. Nat.* **108**, 207-228 (1974).
- Caswell, H. *Am. Nat.* **112**, 127-154 (1978).
- Hastings, A. *Ecology* **68**, 117-123 (1987).
- Hanski, I. *Ecology* **64**, 493-500 (1983).
- Hanski, I. *Oikos* **38**, 210-221 (1982).
- Järvinen, O. & Ulfstrand, A. S. *Oecologia* **46**, 186-195 (1980).
- MacArthur, R. H. & Wilson, E. O. *Evolution* **17**, 373-387 (1963).
- MacArthur, R. H. in *Population Biology and Evolution* (ed. Lewontin, R. C.) 159-176 (Syracuse University Press, 1968).
- Diamond, J. M. & May, R. M. *Science* **197**, 266-270 (1977).

Selective dendritic transport of RNA in hippocampal neurons in culture

Lauren Davis*, Gary A. Banker† & Oswald Steward*‡

* Departments of Neuroscience and Neurosurgery, University of Virginia School of Medicine, Charlottesville, Virginia 22908, USA

† Department of Anatomy, Albany Medical College, Albany, New York 12208, USA

Typical neurons of the central nervous system (CNS) elaborate tens of thousands of membrane specializations at sites of synaptic contacts on their dendrites. To construct, maintain, and modify these specializations, neurons must produce and deliver the appropriate molecular constituents to particular synaptic sites. Previous studies have revealed that polyribosomes are selectively positioned beneath postsynaptic sites^{1,2}, suggesting that in neurons, as in other cell types^{3-8,19}, protein synthetic machinery is located at or near the sites where particular proteins are needed. The mechanisms that deliver ribosomes and messenger RNA to their specific destinations in cells are therefore of considerable interest. Here we describe a system for RNA transport in dendrites that could provide a mechanism for the delivery of ribosomes and mRNA to synaptic sites in dendrites. Hippocampal neurons grown in culture incorporate ³H-uridine in the nucleus, then selectively transport the newly synthesized RNA into dendrites at a rate of about 0.5 mm day⁻¹. The transport is inhibited by metabolic poisons, suggesting that it is an active, energy-dependent process. The RNA may be transported in association with the cytoskeleton.

The present study takes advantage of a unique model system provided by hippocampal neurons grown in culture^{9,10}. These cells develop elaborate dendritic arborizations which in many respects resemble their counterparts *in situ* (reviewed in ref. 11). The dendrites have a characteristic morphology and receive synaptic contacts from other cells. Moreover, the neuron-specific microtubule associated protein MAP2 is selectively localized in the dendrites and somata of these neurons¹²; immunostaining for MAP2 thus provides a convenient method to trace the entire dendritic arbors of individual neurons.

‡ To whom correspondence should be addressed.

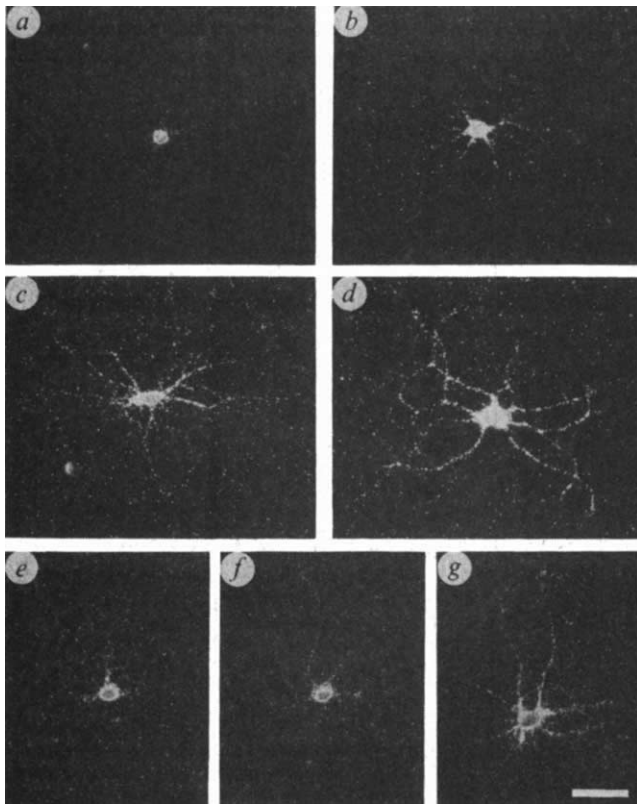


Fig. 1 The migration of ^3H -RNA in dendrites. Autoradiographs of hippocampal neurons incubated for 1 h in ^3H -uridine and fixed at various times after the pulse. Silver grains appear bright in these darkfield micrographs. *a*, Cultures fixed immediately following 1-h pulse. Note that labelling is largely over the nucleus. *b-d*, Cultures transferred to a chase medium containing an excess of unlabelled uridine for 3, 6 and 24 h respectively prior to fixation. *e* and *f*, Inhibition of dendritic transport of RNA by metabolic poisons. Cultures were pulse-labelled for 1 h and then transferred to medium containing azide (10 mM) or dinitrophenol (10 mM), respectively, and an excess of unlabelled uridine for 6 h prior to fixation (compare with *c*). *g*, Association of recently synthesized RNA with the cytoskeleton. Cultures were pulse-labelled and allowed to transport RNA for 24 h (compare with *d*). They were then extracted for 10 min with Triton X-100 (0.3%) in a microtubule stabilizing buffer (0.13 M HEPES, 10 mM EGTA, 20 mM MgSO_4 , pH 6.8) prior to fixation. Scale bar: 50 μm .

Methods. Hippocampal cells were obtained from 18-day embryonic rat brains, plated onto polylysine-treated glass coverslips and maintained in serum-free medium for 15 days as described previously^{9,10}. Cultures were pulse-labelled for 1 h in ^3H -uridine (40 μM Ci ml^{-1}) at 37 $^{\circ}\text{C}$. The chase was carried out in a medium containing 10^{-5} M of unlabelled uridine. Fixation was with paraformaldehyde. For autoradiography, coverslips were mounted cell-side-up on glass slides, and the slides were coated with undiluted NTB2 photographic emulsion (41 $^{\circ}\text{C}$). The slides were exposed for 2–7 days at 4 $^{\circ}\text{C}$. Following the exposure period the cultures were developed in Kodak D-19 (15 $^{\circ}\text{C}$), fixed, and coverslipped with Kaiser glycerol jelly mountant.

Cultures of hippocampal neurons were incubated with ^3H -uridine for 1 h and then either fixed immediately or incubated in unlabelled medium for varying intervals to allow the transport of ^3H -RNA into the dendrites; cultures were then processed for autoradiography. At the end of a 1 h pulse-labelling period (0 h chase interval), label was found only over the cell body; the heaviest labelling was over the nucleus (Fig. 1*a*). With increasing time, ^3H -RNA was found progressively further into the dendrites, reaching the tips of the dendrites within 12 h (Fig. 1*b-d*). When cultures were exposed to an RNA synthesis inhibitor, actinomycin D (10^{-5} M), before and during the ^3H -uridine pulse, they were unlabelled. In cells that were immunostained to localize MAP2 and then processed for autoradiography, the processes that were radioactively labelled invariably were MAP2 positive (Fig. 2); there was never detectable labelling of the thin, MAP2 negative axons.

To estimate the rate of transport, the distance of migration

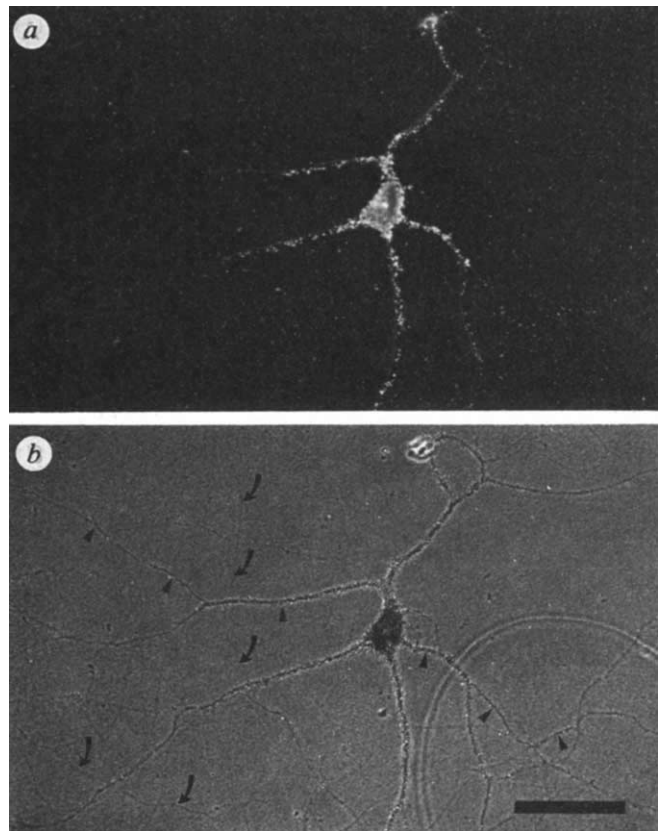


Fig. 2 RNA was selectively transported into dendrites. To establish this, autoradiography was combined with immunostaining for MAP2 to identify the dendrites. Cultures were pulse-labelled with ^3H -uridine and allowed to transport ^3H -RNA for 12 h. *a*, A neuron photographed first by darkfield, then by fluorescence microscopy (double exposure). MAP2-stained dendrites, visualized with a rhodamine-conjugated antibody, appear grey; silver grains appear white. Silver grains are exclusively associated with MAP2-stained dendrites; in some cases the silver grains stop just short of the tips of the dendrites. In preparing this micrograph, the image of the silver grains was shifted slightly to avoid obscuring the fluorescence image. *b*, The same field photographed by phase contrast microscopy illustrating all neuronal processes; arrows = axons, arrowheads = dendrites. Notice that many MAP2 negative axons intersect the MAP2 positive dendrites. There is no evidence of radiolabelling in the axons. Scale bar: 50 μm .

Methods. Cultures which had transported ^3H -RNA for 12 h were fixed and immunoreacted with mouse anti-MAP2 antibody (clone AP-14) followed by rhodamine anti-mouse IgG. They were then processed for autoradiography as described in Fig. 1.

was calculated at each chase interval (Fig. 3*a*). The average distance of transport was roughly linear for the first 6 h and reached a maximum by about 12 h when the average distance of labelling in dendrites approximated the average length of the dendrites based on MAP2 staining. Although there was considerable variability in the distance over which dendrites were labelled (that is, mean length of label after transport for 12 h was 85 μm , s.d. 47), this corresponded to the variability in dendritic length revealed by MAP2 staining (mean length of dendrites = 98 μm , s.d. 58). Because many dendrites are short, labelling that reached the tips of short dendrites at early chase intervals would impose an artificial ceiling on the apparent rate of transport. To provide a more accurate estimate of the rate of migration, we analysed only the longest dendrite of each cell; this analysis indicated a transport rate of about 0.4–0.5 mm day^{-1} (Fig. 3*b*). This is within the range of rates reported for slow axonal transport¹³.

As many forms of transport are energy-dependent¹³, we investigated whether the migration of RNA into dendrites could be blocked by metabolic inhibitors. Cells were pulse-labelled, then treated with dinitrophenol or azide; in treated cells there was essentially no migration of label into the dendrites

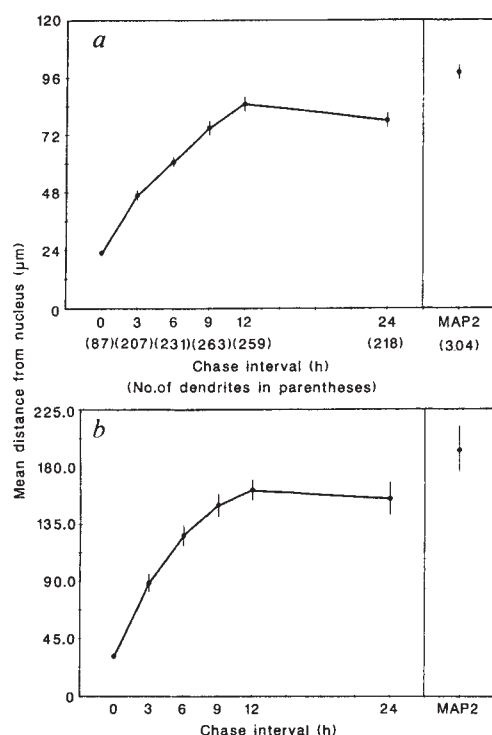


Fig. 3 Rate of migration of RNA into dendrites. The average distance of labelling in dendrites was determined for: *a*, all dendrites of a group of 25 neurons (the total number of dendrites included in each chase interval is listed along the abscissa); *b*, the single dendrite showing the longest distance of labelling in each of 25 neurons (a total of 25 dendrites at each chase interval). The label reached the tips of the dendrites by 12 h. The average length of the dendrites was measured in sister cultures that were immunostained with a MAP2 antibody.

Methods. Twenty-five neurons at each chase interval were randomly chosen on the basis of predetermined microscope stage coordinates and photographed in darkfield illumination. The distance of dendritic labelling was defined by determining the point along each dendrite where grain density decreased to fewer than three silver grains in a template circle 10 μm in diameter. Over proximal dendrites where silver grains were closely packed, up to 50 silver grains fit into a circle of this area; only about 0.3 grains per circle were found over background regions. The distance between the centre of the nucleus and the furthest extent of labelling within each dendrite was determined, and the average distance of transport at each chase interval was calculated.

(Fig. 1*e,f*). At least two-thirds of the neurons remained viable following treatment with these metabolic poisons, based on uptake of fluorescein diacetate¹⁴ (C. Dotti & G. A. B., unpublished observations).

The absence of labelling in axons suggests that RNA cannot move freely throughout the cell. This could be due to binding of mRNA and associated ribosomes to the cytoskeleton, as has been reported to occur in a number of non-neuronal cell types¹⁵. To investigate whether recently synthesized RNA that migrated into dendrites was bound to the cytoskeleton, the cytosolic portion of the cells was extracted by detergent treatment under conditions that preserved assembled cytoskeletal elements. As illustrated in Fig. 1*g*, dendritic labelling was still prominent in the cytoskeletal preparations. The binding of labelled RNA to the cytoskeleton makes it very unlikely that the movement of recently synthesized RNA in dendrites is due to diffusion or other passive processes.

The present results suggest that dendrites have active transport mechanisms that deliver different materials to the axonal transport system. Because recently synthesized RNA present in dendrites is bound to the cytoskeleton, RNA may be transported in conjunction with some component of the dendritic cytoskeleton. It has been suggested that slow axonal transport, which occurs at about the same rate as the dendritic transport of RNA, represents the translocation of assembled cytoskeletal polymers¹⁶. Because the cytoskeleton of dendrites is molecularly

distinct from that of axons^{12,17} both the sorting and the transport of RNA could be a result of its selective binding to the dendritic cytoskeleton. It would be of interest to know if materials also move at more rapid rates in dendrites, comparable to the rate of rapid axonal transport.

In neurons polyribosomes are selectively localized beneath synaptic junctions^{1,2}; in muscle fibres mRNA for the acetylcholine receptor is preferentially localized beneath the neuromuscular junction¹⁸. These findings suggest that the dendritic transport of RNA could play an important role in neuronal function. For example, the transport and selective localization of RNA could enable the synthesis of proteins destined for particular synaptic sites to be regulated locally. It will be of interest to explore whether the transport of RNA is regulated during dendritic growth and synaptogenesis or in conjunction with activity-dependent changes in dendritic excitability.

We thank C. Fitzgerald, who prepared the cell cultures, Drs L. Binder and L. I. Rebhun for a gift of MAP2 antibody. This work was supported by the NIH.

Received 8 July; accepted 7 October 1987.

1. Steward, O. & Levy, W. B. *J. Neurosci.* **2**, 284-291 (1982).
2. Steward, O. in *The Biology of Change in Otolaryngology* (eds Ruben, R. W. et al.) 235-254 (Elsevier, Amsterdam, 1986).
3. Steward, O., Banker, G., Davis, L. & Phillips, L. L. in *Cellular and Molecular Aspects of Neural Development and Regeneration* (eds Gorio, A., deVellis, J., Perez-Polo, J. R. & Haber, B.) (Liss, New York, in the press).
4. Fulton, A. B., Wan, K. M. & Penman, S. *Cell* **20**, 849-857 (1980).
5. Isaacs, W. B. & Fulton, A. B. *Proc. natn. Acad. Sci. U.S.A.* **84**, 6174-6178 (1987).
6. Melton, D. A. *Nature* **328**, 80-82 (1987).
7. Mentaberry, A. et al. *Proc. natn. Acad. Sci. U.S.A.* **83**, 1111-1114 (1986).
8. Lawrence, J. B. & Singer, R. H. *Cell* **45**, 407-415 (1986).
9. Banker, G. A. & Cowan, W. M. *Brain Res.* **126**, 397-425 (1977).
10. Bartlett, W. P. & Banker, G. A. *J. Neurosci.* **4**, 1944-1953 (1984).
11. Banker, G. A. & Waxman, A. B. in *Intrinsic Determinants of Neuronal Form and Function* (eds Lasek, R. J. & Black, M. M.) (Liss, New York, in the press).
12. Caceres, A., Banker, G., Steward, O., Binder, L. & Payne, M. *Dev. Brain Res.* **13**, 314-318 (1984).
13. Grafstein, B. & Forman, D. S. *Physiol. Rev.* **60**, 1167-1283 (1980).
14. Rotman, B. & Papermaster, B. W. *Proc. natn. Acad. Sci. U.S.A.* **55**, 134-141 (1966).
15. Lenk, R., Ransom, L., Kaufmann, Y. & Penman, S. *Cell* **10**, 67-78 (1977).
16. Lasek, R. J. *Phil. Trans. R. Soc. B299*, 313-327 (1982).
17. Shaw, G., Banker, G. A. & Weber, K. *Eur. J. Cell Biol.* **39**, 205-216 (1985).
18. Merlie, J. P. & Sanes, J. R. *Nature* **317**, 66-68 (1985).
19. Black, M. M., Keyser, P. & Sobel, E. *J. Neurosci.* **6**, 1004-1012 (1986).

A novel epitope of the LFA-1 antigen which can distinguish killer effector and suppressor cells in human CD8 cells

Chikao Morimoto, Christopher E. Rudd,
Norman L. Letvin & Stuart F. Schlossman

Division of Tumor Immunology, Dana-Farber Cancer Institute,
Department of Medicine, Harvard Medical School, Boston,
Massachusetts 02115, USA

The CD4 subset of cells displays helper/inducer activity and recognizes class II antigens of the major histocompatibility complex (MHC), while the CD8 subset recognizes class I MHC antigens and exhibits cytotoxic or suppressor function¹⁻³. Considerable functional as well as corresponding phenotypic heterogeneity exists within the two major T cell subsets⁴⁻⁸. Although the CD8⁺ population contains pre-cytotoxic, cytotoxic, pre-suppressor and suppressor effector T cells, these distinctions still rest largely on the use of functional assays. Attempts have been made to define the CD8⁺ precursor of the killer cell with new monoclonal antibodies⁹. But more precise phenotypic distinctions between the functional subpopulations within CD8⁺ cells will be needed. We have now developed a monoclonal antibody, anti-S6F1 which can distinguish killer effector and suppressor effector cells in CD8 lymphocyte populations. The cell-surface structure defined by this antibody comprises two glycoproteins with relative molecular mass (*M_r*)

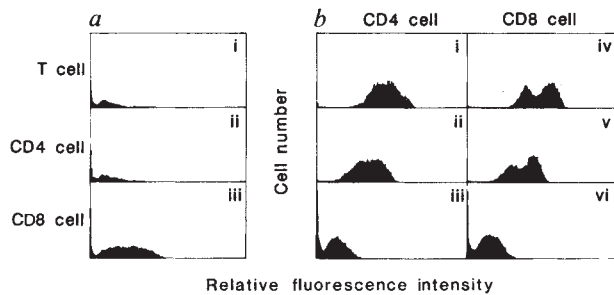


Fig. 1 *a*, Cytofluorographic analysis of unfractionated T (i), CD4 (ii) and CD8 (iii) cells with anti-S6F1 monoclonal antibody. Each histogram shows the number of cells (ordinate) against fluorescence intensity (abscissa). The mean percent reactivity $\pm$ s.e.m. of anti-S6F1 antibody with lymphocytes from ten individuals were as follows: S6F1 was expressed on $10 \pm 2\%$ of unfractionated T cells, $14 \pm 2\%$ of CD4⁺ cells and $49 \pm 3\%$ of CD8⁺ cells. *b*, Each histogram shows the number of cells (ordinate) against fluorescence intensity (abscissa).

Methods. *a*, The monoclonal antibody anti-S6F1 (IgG1 subclass) was produced by standard techniques after immunizing a BALB/c mouse with the line 1670, an immortalized splenocyte population derived from a *Herpesvirus saimiri* strain 11 infected whitelip tamarin. PBL were obtained from healthy volunteer donors by Ficoll-Hypaque density gradient centrifugation and PBL were separated into T and non-T cells by rosette formation with sheep erythrocytes as described previously². CD4⁺ and CD8⁺ cells were obtained by complement-mediated lysis with anti-CD8 and anti-CD4 respectively⁴. 1×10^6 cells were analysed for their reactivity with anti-S6F1 antibody (1:1,000) by indirect immunofluorescence on an Epics C cell sorter (Coulter). *b*, The reactivity of CD4 and CD8 cells with varied dilution of anti-LFA-1 antibody (2F12) were analysed by indirect immunofluorescence on an Epics C cell sorter. i, iv, anti-LFA-1 antibody 1:100 (98% cell reactivity); ii, v, anti-LFA-1 antibody 1:1,000 (98% reactivity); iii, vi, anti-LFA-1 antibody 1:10,000 (52% of CD4 cell and 56% of CD8 reactivity).

180K and 95K respectively. Also sequential immunoprecipitation studies and two dimensional gel electrophoresis indicate that anti-S6F1 recognizes a novel epitope on the LFA-1 antigen¹⁰⁻¹².

Figure 1a shows the fluorescence profile of the expression of S6F1 antigen on unfractionated T lymphocytes, CD4⁺ lymphocytes and CD8⁺ lymphocytes. In this representative study, anti-S6F1 was reactive with 17% of unfractionated T cells, only 14% of CD4⁺ cells and 58% of CD8⁺ cells. Thus, the S6F1 antigen was expressed preferentially on the CD8⁺ subpopulation of lymphocytes. Anti-S6F1 reacted with greater than 70% of null cells, granulocytes and the T-cell line HSB. But, it reacted with only a small percentage (<10%) of cells in populations of macrophages, B cells, B-cell lines (Raji, Ramos, Nalm-1, Epstein-Barr virus transformed cell lines), haematopoietic lines (U-937, K-562, KG-1) or T-cell lines (Molt 4, CEM, JM) (data not shown).

Because anti-S6F1 reacts with a sizeable population of CD8⁺ cells, we used this antibody to subdivide the CD8⁺ population of peripheral blood lymphocytes (PBL) to assess the functional heterogeneity of these cells. We investigated whether alloreactive cytotoxic T lymphocytes, either precursor or effector cells, can be separated from CD8⁺ suppressor cells with the anti-S6F1 antibody. To determine whether the precursor of alloantigen-specific cytotoxic T cells belongs to either the S6F1⁺ or S6F1⁻ population of CD8⁺ cells, we examined the specific cell-mediated lympholysis (CML) by subsets of CD8⁺ cells. As shown in Fig. 2a, the majority of precursors of cytotoxic T cells against specific target cells resided in the CD8⁺S6F1⁻ subset, while a small proportion was found in the CD8⁺S6F1⁺ subset.

We then determined whether the cytotoxic effector cell belonged to either the mixed lymphocyte reaction (MLR)-activated CD8⁺S6F1⁺ or CD8⁺S6F1⁻ lymphocyte subpopulation. As shown in Fig. 2b and in contrast to the results shown in Fig. 2a, the CD8⁺S6F1⁺ lymphocyte population demonstrated

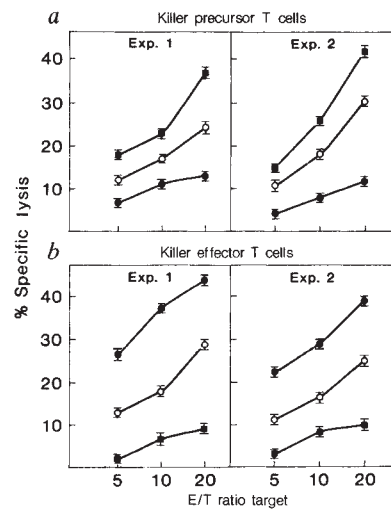


Fig. 2 *a*, Killer precursor T cells belonged to CD8⁺S6F1⁻ cells. CD8 control (○-○), CD8⁺S6F1⁺ (●-●) and CD8⁺S6F1⁻ (■-■). *b*, Killer effector T cells belonged to CD8⁺S6F1⁺ cells.

Method. Freshly isolated CD8⁺ cells were labelled with anti-S6F1 antibody and developed with fluorescein-conjugated F(ab')₂ goat anti-mouse F(ab')₂ (Tago). By using an Epics C cell sorter CD8 cells were separated into CD8⁺S6F1⁺ and CD8⁺S6F1⁻ populations as described previously⁴. Then 1×10^6 ml⁻¹ of CD4⁺ cells obtained by complement-mediated lysis added to generate killer inducer function in this system were cultured with 1×10^6 ml⁻¹ of unfractionated CD8⁺, sorted CD8⁺S6F1⁺ or CD8⁺S6F1⁻ cells in the presence of equal numbers of irradiated allogeneic stimulator cells (E⁻ cells) in the 24 wells culture plates (Costar) in a 5% CO₂ humidified atmosphere at 37 °C. At the end of six days of culture, after removing CD4 cells from cultures by complement-mediated lysis of CD4⁺ cells, ⁵¹Cr release cell-mediated lympholysis by subsets of CD8 cells was assayed after a 4-h cell incubation as described previously². *b*, Unfractionated T cells (1×10^6 ml⁻¹) were sensitized with equal numbers of irradiated (5,000 Rad) allogeneic (E⁻) stimulator cells in final RPMI 1640 medium (10% pooled human AB serum, (Pel Freeze), 4 mM L-glutamine, 25 mM HEPES buffer (Microbiological Associates), 0.5% sodium bicarbonate and 10% penicillin-streptomycin) in a 25 cm² flask in 5% CO₂ humidified atmosphere at 37 °C for six days. Then the unfractionated, sensitized T cells were separated into CD8⁺S6F1⁺ and CD8⁺S6F1⁻ subset of cells on an Epics C cell sorter. Both unfractionated and fractionated CD8⁺ cells were analysed in cell-mediated lympholysis against ⁵¹Cr-labelled targets. Cryopreserved allogeneic (E⁻) cells used as targets were thawed the previous day, cultured overnight and labelled with ⁵¹Cr by incubation for one hour at 37 °C in 0.2 ml of sodium chromate in saline (1 m Ci ml⁻¹, New England Nuclear). Targets were then washed and resuspended at 10^5 ml⁻¹ in final media. A standard 4-h cytotoxicity assay was performed as previously described using V-bottom wells.

greater allospecific killer effector activity than did the unfractionated CD8⁺ population. Furthermore, virtually no killer effector activity was observed in the CD8⁺S6F1⁻ lymphocyte population. Although anti-MHC class I and anti-LFA-1 antibodies blocked the killer effector function as previously described^{2,11,12}, anti-S6F1 did not block the effector function (data not shown). These results clearly show that killer precursor T cells are found in the CD8⁺S6F1⁻ subset of cells but the killer effector activity belongs to the CD8⁺S6F1⁺ subset.

The above results suggested that the S6F1 antigen can become expressed on a S6F1 negative population of CD8⁺ cells during a primary mixed lymphocyte reaction. To confirm this a primary mixed lymphocyte reaction was initiated using CD8⁺S6F1⁻ cells combined with CD4⁺ cells. These T lymphocytes were then evaluated for S6F1 antigen expression. In three separate experiments, 45-55% of CD8⁺S6F1⁻ cells expressed S6F1 antigen on their cell surfaces after MLR activation. These results clearly indicated that S6F1 antigen expression is induced on a S6F1⁻ population of CD8⁺ cells.

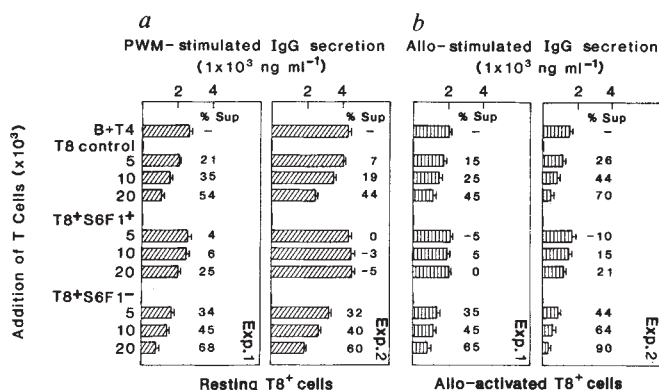


Fig. 3 *a*, Suppressor function of subsets of CD8⁺ cells in a PWM-driven IgG synthesis system. *b*, Suppressor effector function of alloantigen activated subsets of CD8⁺ cells in an alloantigen driven IgG synthesis system.

Method. Freshly isolated unfractionated T cells were separated into CD8⁺S6F1⁺ and CD8⁺S6F1⁻ subset of cells on an Epics cell sorter. Varying numbers of subsets of CD8 cells were added to a constant number of B cells (5×10^4) and CD4⁺ cells (2×10^4) stimulated with PWM in round bottomed 96 well micro-culture plates and total IgG production was measured after seven days in culture. Culture supernatants for IgG secretion were determined by RIA as described previously⁴. *b*, T cells were activated in MLR against alloantigen for six days and alloactivated T cells were separated into CD8⁺S6F1⁺ and CD8⁺S6F1⁻ cells on an Epics C cell sorter. Unseparated CD8⁺ and subsets of CD8⁺ cells were added to autologous PBL in the presence of irradiated allogeneic cells and the effects on alloantigen-driven IgG synthesis were assessed. Percent suppression of IgG synthesis was calculated as follows:

$$\% \text{ suppression} = 1 - \frac{\text{observed IgG}}{\text{IgG in no regulator cells}} \times 100$$

We next investigated which subset of CD8⁺ cells was activated to become suppressor effector cells in a pokeweed mitogen (PWM)-driven IgG synthesis system. As shown in Fig. 3*a*, the addition of increasing numbers of unfractionated CD8⁺ cells resulted in the suppression of PWM driven IgG secretion by B cells in a dose-dependent fashion. When increasing numbers of CD8⁺S6F1⁻ cells were added to CD4⁺ and B cells, an even greater degree of suppression of IgG secretion was observed. In contrast, the addition of CD8⁺S6F1⁺ cells to the mixture of B and CD4⁺ cells resulted in only a slight decrease in IgG secretion. These results show that the precursor of CD8 suppressor cells belongs to the CD8⁺S6F1⁻ subset of cells.

To determine whether suppressor effector activity was found in the CD8⁺S6F1⁺ population, we examined the suppressor function of the subsets of CD8⁺ cells after MLR activation in an alloantigen-stimulated IgG synthesis system. As shown in Fig. 3*b*, alloactivated unfractionated CD8⁺ cells suppressed alloantigen-driven IgG synthesis in a dose-dependent manner. The CD8⁺S6F1⁻ population activated with alloantigen showed greater suppressor activity in this system than did the unfractionated CD8⁺ cells. But alloantigen-activated CD8⁺S6F1⁺ lymphocytes exhibited a minimal suppressor activity only when very large numbers of cells were added to these cultures. These results indicate that both suppressor precursor and suppressor effector cells belong to the CD8⁺S6F1⁻ lymphocyte population. More importantly, the above results imply that suppression is probably not merely a manifestation of cytotoxicity.

Given the importance of the S6F1 antigen in the identification of cytotoxic effector T cells, we analysed the structure of the antigen. Figure 4 shows a composite pattern of polypeptides precipitated by the anti-S6F1 antibody and the anti-2F12 antibody (anti-LFA-1) (ref. 10) from ConA activated T cells. Interestingly, the anti-S6F1 antibody precipitated two major bands corresponding to 180K and 95K (Fig. 4*a*, lane 1); a pattern

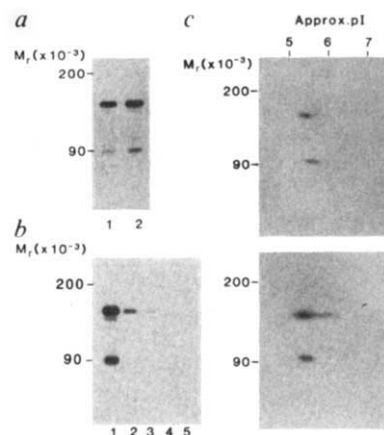


Fig. 4 Immunoprecipitation analysis of the S6F1 and LFA-1 antigens. *a*, Lane 1, anti-S6F1 antibody; lane 2, 2F12 antibody. *b*, Represents the sequential immunodepletion of the S6F1 and LFA-1 antigens. Lysates from labelled peripheral blood lymphocytes were precleared with the TS1/18 antibody for four immunoprecipitations (*b*, lanes 1–4), prior to precipitation with the S6F1 antibody (*b*, lane 5). *c*, Shows the two-dimensional non-equilibrium focusing (NEPHGE/SDS-PAGE) analysis of the S6F1 and LFA-1 antigens. Figure 4*c*: upper panel, the S6F1 antibody; lower panel, the 2F12 antibody.

Method. Cells from T cell-enriched cells (>95% E⁺ cells) were activated with ConA ($20 \mu\text{g ml}^{-1}$) for 3 days and were labelled with ¹²⁵I by lactoperoxidase-catalysed iodination¹³, washed and then solubilized in lysis buffer (1% w/v Nonidet P-40 in 20 mM Tris-HCl buffer pH 8.0 containing 150 mM NaCl, 1 mM EDTA and 1 mM phenylmethyl (sulphonyl) fluoride). The lysate (5×10^7 cell equivalents) was then incubated overnight at 4 °C with 5 μl of ascitic fluid of the individual antibodies before precipitating with 50 μl of 10% w/v Protein A: Sepharose. The immunoprecipitates were then washed once in lysis buffer with 0.1% (w/v) SDS and twice with lysis buffer alone, prior to analysis by SDS-PAGE (ref. 14). Lysates were labelled for 2-D gel analysis, in which the immune complexes were eluted from the beads by incubation in isoelectric focusing sample buffer (9.2 M urea, 2% Nonidet P-40, 2% ampholines pH range 2–11 (Serva) at 50 °C for 30 min. Two dimension NEPHGE/SDS-PAGE analysis was done as previously described^{15,16}.

which appear to correspond to the α and β subunits of the LFA-1 antigen, respectively (Fig. 4*a*, lane 2).

A direct demonstration of the identity of the S6F1 antigen was made by sequential immunodepletion and two dimensional non-equilibrium gel electrophoresis (2-D NEPHGE/SDS-PAGE). As shown in Fig. 4*b*, monoclonal antibody against the LFA-1 antigen was found to deplete completely material reactive with the S6F1 antibody. The monoclonal antibody TS1/18 has previously been shown to react with the β subset of the LFA-1 structure^{11,12}. Also, analysis by 2-D NEPHGE/SDS-PAGE showed that antibodies against the S6F1 and LFA-1 antigens precipitated two polypeptides of 180K and 95K which focused at approximate isoelectric positions of pH 5.3 and 5.7, respectively (Fig. 4*c*). These data suggest that the anti-S6F1 antibody reacts with a LFA-1 antigen.

While anti-S6F1 and anti-2F12 immunoprecipitate the same structure, their reactivity with functionally distinct cell subsets appears to differ. To rule out the possibility that these differences were not merely a manifestation of differing antibody affinities for the LFA-1 structure, we examined the reactivity of both CD4⁺ and CD8⁺ lymphocyte populations with various dilutions of anti-2F12. As shown in Fig. 1*a* and *b*, while varying dilutions of anti-2F12 from $1:10^2$ to $1:10^4$ stained 52% (Fig. 1*b*-iii) and 56% (Fig. 1*b*-iv) of CD4⁺ and CD8⁺ lymphocytes, respectively, whereas anti-S6F1 ($1:10^3$) stained CD8⁺ lymphocytes preferentially. Thus the reactivity of CD8⁺ lymphocytes with anti-S6F1 does not appear to reflect simply a low affinity of the antibody for the LFA-1 structure itself.

The actual epitope recognized by the anti-S6F1 antibody appears not to be involved in the process of cytotoxicity as shown by our inability to block this function with anti-S6F1 (data not shown) whereas other antibodies directed against the LFA-1 antigen have been shown effectively to block these events^{11,12}. Moreover, although the LFA-1 antigen is expressed on a broad spectrum of cell types^{11,12}, the S6F1 antigen is expressed primarily on a subset of T cells, null cells and a small fraction of macrophages and B cells. It is possible that the S6F1 epitope may be formed distantly on the molecule by the development of a binding site on the LFA-1 antigen involved in adherence to target cells.

We thank Ms M. Hagan and Ms M. Dimaggio for technical assistance, Mr Herbert Levine for his flow cytometric analysis and Dr T. Springer for his generous gift of anti-TS1/18 antibody. We also thank Ms I. Isherwood for secretarial assistance. This work was supported in part by the NIH and the Cancer Research Institute.

Received 20 August; accepted 12 October 1987.

- Reinherz, E. L. & Schlossman, S. F. *New Engl. J. Med.* **303**, 370-373 (1980).
- Meuer, S. C., Schlossman, S. F. & Reinherz, E. L. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4395-4399 (1982).
- Krensky, A. M. *et al. J. Immun.* **129**, 2001-2003 (1982).
- Morimoto, C., Letvin, N. L., Distaso, J. A., Aldrich, W. R. & Schlossman, S. F. *J. Immun.* **134**, 1508-1515 (1985).
- Morimoto, C. *et al. J. Immun.* **134**, 3762-3765 (1985).
- Smith, S. H., Brown, M. H., Rowe, D., Callanard, R. E. & Beverley, P. C. L. *Immunology* **58**, 63-70 (1986).
- Reinherz, E. L., Hussey, R. E. & Schlossman, S. F. *Immunogenetics* **11**, 421-426 (1980).
- Gatenby, P. A., Kotzin, B. L., Kansas, G. S. & Engleman, E. G. *J. exp. Med.* **156**, 55-67 (1982).
- Clement, L. T., Grossi, C. E. & Gartland, G. L. *J. Immun.* **133**, 2461-2468 (1984).
- Tedder, T. F., Schmidt, R. E., Rudd, C. E., Kornacki, M. M., Ritz, J. & Schlossman, S. F. *Eur. J. Immun.* **16**, 1539-1543 (1986).
- Sanchez-Madrid, F., Nagy, J., Robbins, E., Simon, P. & Springer, T. A. *J. exp. Med.* **158**, 1785-1803 (1983).
- Sanchez-Madrid, F. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 7489-7493 (1982).
- Hubbard, A. L. & Cohn, Z. A. *Biochemical analysis of membranes* (ed. Maddy, A. H.) 427-501 (Chapman and Hall, London, 1976).
- Laemmli, U. K. *Nature* **267**, 680-684 (1970).
- O'Farrell, P. L., Goldman, H. J. M. & O'Farrell, P. H. *Cell* **12**, 1133-1142 (1977).
- Rudd, C. E., Bodmer, J. G., Bodmer, W. F. & Crumpton, M. J. *J. biol. Chem.* **260**, 1927-1936 (1985).

Expression of immunoglobulin delta chain causes allelic exclusion in transgenic mice

Antonio Iglesias, Marinus Lamers & Georges Köhler

Max-Planck-Institut für Immunbiologie, Stübeweg 51, D-7800 Freiburg, FRG

Rearrangement of the separated V, D and J segments of immunoglobulin heavy and light chain genes is an ordered and regulated process essential for the production and diversification of antibodies in mammals¹. Only one allele of the heavy and light chain gene locus is found functionally rearranged in normal B cells. The second allele is either non-functionally or incompletely rearranged, a phenomenon known as allelic exclusion. Recently, we² and others³ have shown that expression of a rearranged μ gene introduced into the germline of mice leads to inhibition of rearrangement of endogenous heavy chain genes. It has been suggested that μ also exerts a positive influence on kappa light chain gene rearrangement⁴. Here we show that a functionally rearranged δ gene is also able to prevent endogenous heavy chain gene rearrangement and could promote κ light chain gene rearrangement.

Immunoglobulin D (IgD) is transiently expressed on the surface of mature, non-stimulated B cells together with IgM (see ref. 5 for review). After antigen stimulation, production of IgD ceases. In the majority of cells this is followed by expression of other immunoglobulin isotypes, a process called immunoglobulin class switch⁶. Isotype switching is accompanied

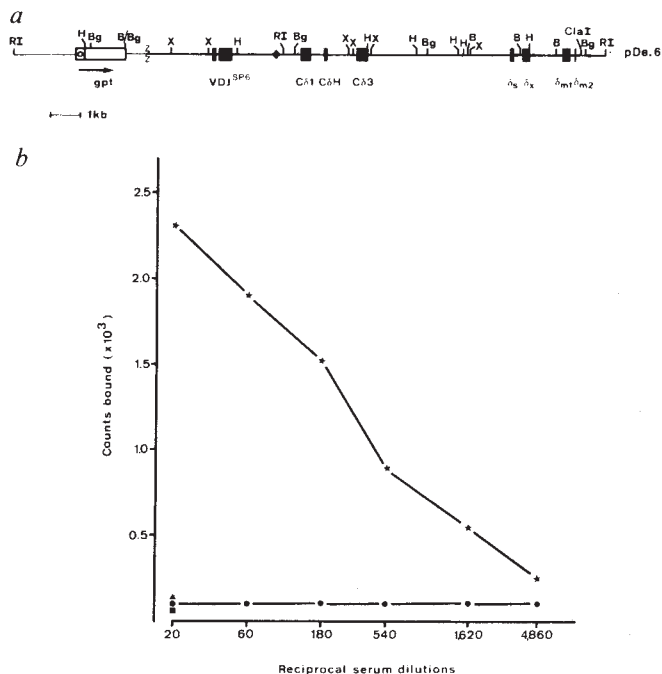


Fig. 1 *a*, Restriction map of pDe.6. Mouse sequences are indicated as thick lines. The corresponding δ -exons (black boxes) are designated as in ref. 5. $\blacklozenge$, Denotes the immunoglobulin heavy chain enhancer; $\bullet$ indicates positions of polyadenylation sites. The pSV2-gpt vector sequences are shown as thin lines. The *Escherichia coli* gpt gene, together with the SV40 origin of replication ($\circ$) is depicted as an empty box. The arrow indicates the direction of transcription of gpt which is the same as for the δ gene. The vertical wavy line points to the junction between vector and mouse sequences resulting from the deletion detailed below. Restriction enzymes are indicated at their corresponding positions as abbreviations: RI, *Eco*RI; H, *Hind*III; Bg, *Bgl*II; B, *Bam*HI; X, *Xba*I; B/Bg indicates a *Bam*HI-*Bgl*II fusion site. For vector construction, a 5-kilobase (kb) *Eco*RI fragment containing the rearranged VDJ_H gene segment of the μ gene of Sp6, that includes about 3 kb of upstream and 1.6 kb downstream (thus retaining the heavy chain enhancer), has been first inserted into the *Eco*RI site of the vector pSV2-gpt (ref. 21). A deletion of about 0.8 kb has been introduced at the *Eco*RI located 5' of the VDJ_H region (indicated as a vertical wavy line). The remaining unique *Eco*RI site downstream of the enhancer was used to insert a 10-kb *Eco*RI fragment containing the entire C δ region isolated from the phage clone ME645 (ref. 22) (gift from A. Traunecker). The coding capacity of pDe.6 was confirmed by gene transfer and stable expression of δ in the hybridoma Sp6 (data not shown). *b*, IgD^a production in transgenic mice. Serial dilutions of serum of transgenic (*, $\blacktriangle$) or normal siblings ($\bullet$, $\blacksquare$) were incubated in polyvinyl wells previously coated with the monoclonal antibodies 10.4.22 (*, $\bullet$) or 11.6.3 ($\blacktriangle$, $\blacksquare$). These antibodies are specific for the IgD allotypes *a* and *b* respectively¹⁶. The bound material was identified with ¹²⁵I-labelled, anti-Sp6-idiotypic antibody 20-5 (P. A. Cazenave, unpublished).

by excision of constant region μ and δ sequences in the heavy chain gene locus⁷. The time course of IgD expression at the onset of an immune response has suggested a possible function for IgD in immune phenomena such as immunoglobulin class switch, the induction of immune memory⁵ and of helper T cells⁸, but no precise biological function for IgD has been found. To this end, we have chosen a transgenic mouse model⁹ for our studies of IgD. The rearranged VDJ gene segment of the μ gene of the Sp6 hybridoma¹⁰ (including the immunoglobulin heavy chain gene enhancer) was joined to the genomic δ constant region of BALB/c (allotype *a*, see Fig. 1a for details). About 300-500 copies of the linearized plasmid pDe.6 have been injected into fertilized (C57B1/6 $\times$ SJL)F1 eggs. This strain combination carries the IgD allotype *b*, and allows easy discrimination between the transgenic and endogenous δ chains. Furthermore,

Table 1 Flow-cytometric analysis (spleen)

Cell-surface determinant	Littermate control	Transgenic
δ^a positive	0.2	30.5
μ positive	50.2	11.0
δ^b positive	37.6	6.5
Idiotypic positive	0.0	24.4
κ positive	52.4	32.4
μ positive, δ^a positive	1.9	7.1
CD-4 positive	15.0	18.8
CD-8 positive	10.3	12.7

Pooled spleen cells of two transgenic mice and two littermates were stained with fluoresceinated or biotinylated monoclonal antibodies to IgD^a (clone 10.4.22) (ref. 16), IgD^b (clone 11.6.3) (ref. 16), IgM (clone 331.12) (ref. 17), Sp6 idiotype (clone 20-5) (Cazenave, unpublished), kappa (clone 187-1) (ref. 18), CD-4 (clone GK1.5) (ref. 19) or CD-8 (clone 53-6.7) (ref. 20) followed, in the case of biotinylated antibodies by a second incubation with Texas Red-Streptavidin (Amersham). All staining and washing procedures were performed at 4 °C in the presence of 0.1% sodium azide. Cells were analysed on an Ortho double-laser Cytofluorograph, results are given as the percentage of cells positive for the surface determinant analysed. The average number of nucleated cells in the spleen of transgenic mice was 4×10^7 , in the littermate it was 8×10^7 cells.

the Sp6 heavy chain variable region is recognized by the monoclonal antibody 20-5 when associated with most kappa light chains (our unpublished observations).

Four transgenic mice have been generated with copy numbers varying from 3 to 20. Two mice did not have any detectable levels of transgenic IgD in serum. The other two mice had transgenic IgD in serum, but one was infertile. The other mouse (B δ .Sp6-M3, male) had 10–20 copies of the transgene, transmitted the transgene to its progeny and showed large amounts of transgenic protein in the serum. As shown in Fig. 1b, Sp6-idiotype-bearing IgD of the α -allotype is easily detected in the serum of a transgenic offspring, but not in the serum of a normal sibling. Cytofluorometric analysis showed that the transgenic IgD is present on the surface of about 30% of spleen cells of a transgenic mouse, but is not found on spleen cells of a normal littermate (see Table 1). Further, the proportion of surface IgM-positive spleen cells in the transgenic mouse was substantially diminished (about 10% versus 45–50% of the spleen cells of normal littermates, see Table 1). As the spleen size of transgenic animals is about half that of the littermate control (Table 1, legend), the total number of cells per spleen which express endogenous heavy chain genes is reduced by a factor of 10 when compared to the littermate. From the numbers of δ^a - and κ -bearing cells (30% and 32.4%), we conclude that most B cells in the spleen express the transgene. Endogenous IgD^b was detected on 75% of the μ -positive spleen cells of the normal littermate and on 60% of the μ -positive cells in the transgenic

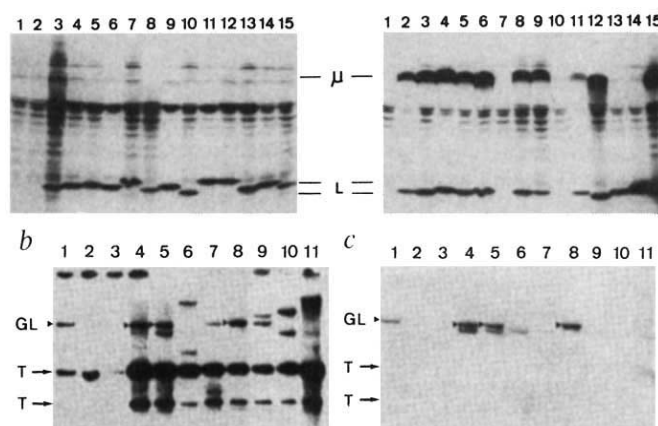


Fig. 2 Analysis of transgenic hybridomas. *a*, Reducing SDS-PAGE of ¹⁴C-labelled culture supernatants of single hybridomas from either a transgenic mouse or a normal littermate. The numbers refer to individual hybridomas of each series. The position of μ and light chains (L) is indicated. Heavy chains have been found in some transgenic hybridomas (not shown, see Table 2). *b*, Representative Southern blot analysis of *Pvu*II-digested DNA of liver from a Sp6-transgenic mouse² (1), the myeloma used for fusion X63Ag8.653(2), and hybridomas made from two transgenic siblings (3–7) and (8–11) respectively. The probe used here is a 2 kb *Bam*HI/*Eco*RI genomic fragment including the J_H3/J_H4 segments and downstream sequences. Germ-line bands (GL) are indicated by arrowheads. Bands contributed by the transgene are also indicated (T). *c*, Same blot as in *b*, after washing and rehybridization with the 0.5-kb *Bam*HI/*Bst*EII genomic DNA fragment immediately 5' upstream of the J_H1 segment. Hybridization with this probe is diagnostic of unrearranged loci. As expected, this probe does not hybridize with transgenic fragments (T). The lower band occasionally seen is due to trace contamination of the probe with pBR sequences. The DNA of lanes 8, 9, 10, 11 in *b* and *c* correspond to the hybridomas of lanes 14, 12, 10 and 8 of Fig. 2a, left, respectively. A mixture of two D_H-specific probes (called DSp2 and DF1-16) (ref. 13) has also been used with the same blot to score for D_H rearrangements (not shown). All DNA probes have been oligolabelled to a specific activity of $>10^9$ c.p.m. μ g⁻¹ of DNA. Hybridizations were done at 65 °C for 12–18 h as described². Autoradiographs were exposed for 4–14 h.

mouse. The distribution of T cells in CD-4 (helper) and CD-8 (suppressor/cytotoxic) positive cells was not changed in the transgenic spleen (Table 1).

To characterize further the B-cell defect, hybridomas from lipopolysaccharide-stimulated spleen cells of two transgenic and one normal littermate have been analysed. Figure 2a shows a comparison of the immunoglobulins produced by 15 hybridomas from a transgenic descendant (Fig. 2a, left) and a non-transgenic littermate (Fig. 2a, right). IgD^a was detected by radioimmuno-

Table 2 Analysis of hybridomas

Mice	Endogenous immunoglobulin production*			Total hybridomas analysed	Rearrangement status†			Total alleles analysed
	H + L	L	0		GL	DJ	VDJ	
δ -transgenic	5	19	4	28	8 $\pm$ (22%)	23 (64%)	5 (14%)	36 (100%)
Normal littermate	18	5	7	30	0	2 (14%)	12 (86%)	14 (100%)

* As detected by SDS-PAGE analysis of hybridoma culture supernatants radiolabelled with ¹⁴C-leucine¹⁰. H, heavy chain; L, light chain; 0, no immunoglobulin chains.

† Transgenic and normal littermate hybridomas were scored for the type of endogenous rearrangement using the Southern blot hybridization technique (see Fig. 2b and c). Included are all data obtained with the three different probes described in Fig. 2 legend. The rearrangement status is indicated: GL, germ-line, non-rearranged; DJ, incomplete rearrangement; VDJ, functional and non-functional complete rearrangement.

‡ Judged by the intensity of the germ-line and the absence of additional rearranged bands two lines had both J_H alleles in the unrearranged germ-line configuration. No GL/VDJ combination has been observed.

assay in 9 of 19 hybridomas which retained the transgene. Although endogenous heavy chains were sometimes found (Table 2), the majority of the transgenic hybridomas produced light chains alone, or in association with transgenic δ (Fig. 2a and Table 2). The reason for the absence of endogenous heavy chains in transgenic hybridomas became apparent when the rearrangement status of the corresponding gene locus was analysed. A Southern blot analysis of DNA from transgenic hybridomas is shown in Fig. 2b and c. Digestion with *PvuII* and hybridization with a probe specific for the genomic region spanning the 3' portion of the mouse J_H -cluster permits the detection of rearrangements (see Fig. 2 legend). A band at a position identical to the non-rearranged germline band (GL) of liver DNA was seen with DNA from several of the transgenic hybridomas (Fig. 2b). Rehybridization with a probe specific for the region immediately 5' to the J_H -cluster, which is lost upon rearrangement (see Fig. 2 legend), confirmed that some of these bands were true germ-line fragments (Fig. 2c). To discriminate between incompletely (DJ_H) and fully (VDJ_H) rearranged DNA configurations, the filters were rehybridized with a mixture of D_H -specific probes (see Fig. 2 legend). Single restriction fragments hybridizing with both the D_H and J_H 3' probes were scored as DJ_H rearrangements (gels not shown). The rearrangement characteristics of all hybridomas analysed is compiled in Table 2. The most striking difference between transgenic and non-transgenic littermates is the high proportion of germ-line alleles (22%) in hybridomas of δ -transgenic mice. The percentage of partially rearranged alleles (of the DJ_H -type) is also notably increased in transgenic hybridomas (64% versus 14% for normal littermates). We conclude from these results that the transgenic δ protein exerts an inhibitory influence on the rearrangement of the endogenous heavy chain genes in the B cells of transgenic mice. This effect is similar to or stronger than that observed in mice transgenic for $\mu + \kappa$ (ref. 3) or μ alone.

It could be argued that the transgenic enhancer, present in about 20 copies, competes with the endogenous enhancer for binding proteins, and thus inhibits expression of the endogenous genes. A μ -transgenic mouse line with 67 copies of the transgene has been described. In this mouse line only the secreted form of μ (μ_s) was made. No inhibition of endogenous μ production was observed (ref. 11 and U. Storb, personal communication). Furthermore, high amounts of messenger RNA and of μ_s protein were made. These results argue against the possibility that endogenous rearrangements are influenced by transgene copy number or mRNA levels. They also show that the secretory form of the μ protein is not sufficient for inhibition of rearrangement. It has been shown that the expression of a human μ gene directing the synthesis of the membrane form exclusively was sufficient for inhibition of rearrangement of endogenous heavy chain genes in a transgenic mouse¹².

The suggestion that μ is a positive signal for κ -gene rearrangement⁴ could also be valid for δ chains, as many of our δ -transgenic hybridomas produce light chains in the absence of functionally rearranged endogenous heavy chain genes (Fig. 2a), and also as 75% of the splenic B cells are surface IgD^a and κ -positive, even though they express no detectable endogenous heavy chains (Table 1). Recently it has been shown that the induction of κ -gene rearrangement could be related to the presence of the membrane form of μ (ref. 13). Our findings suggest that the transmembrane and cytoplasmic domain of μ and probably also of the δ chain are required for allelic exclusion. These domains are very similar and have an unusual short cytoplasmic tail (Lys, Val, Lys). The function of other domains of the μ constant chain in the regulation of the rearrangement process remains to be assessed; domains 2 and 4 seem to be dispensable because the mouse δ -chain lacks the corresponding domains⁵.

The function of δ chains in the regulation of allelic exclusion is surprising because δ is thought to be expressed only after rearrangement and expression of μ and κ genes^{14,15}. Therefore,

the biological significance of IgD for normal B-cell function has been investigated in the generation and regulation of an immune response^{5,8}. Our data relate δ to regulatory processes in the development of B cells, where its function is similar and perhaps complementary to that of μ in the control of immunoglobulin gene rearrangement.

We are grateful to Angelika Lanzendörfer, Annegret Schulz, Carla Jacot and Pascale Renard for technical help. We thank Peter Nielsen and Hermann Eibel for critical reading of the manuscript and Gabriele Eichhorn for typing it.

Received 10 September; accepted 14 October 1987.

1. Tonegawa, S. *Nature* **302**, 575-581 (1983).
2. Rusconi, S. & Köhler, G. *Nature* **314**, 330-334 (1985).
3. Weaver, D., Constantini, F., Imanishi-Kari, T. & Baltimore, D. *Cell* **42**, 117-127 (1985).
4. Reth, M., Ammirati, P., Jackson, S. & Alt, F. *Nature* **317**, 353-355 (1985).
5. Blattner, F. R. & Tucker, P. W. *Nature* **307**, 417-422 (1984).
6. Shimizu, A. & Honjo, T. *Cell* **36**, 801-803 (1984).
7. Radbruch, A. & Sablitzky, F. *EMBO J.* **2**, 1929 (1983).
8. Coico, R. F. *et al. Nature* **316**, 744-746 (1985).
9. Palmiter, R. D. & Brinster, R. L. *Rev. Genet.* **20**, 465-499 (1986).
10. Köhler, G. & Milstein, C. *Eur. J. Immun.* **6**, 511-519 (1976).
11. Storb, U. *et al. J. exp. Med.* **164**, 627-641 (1986).
12. Nussenzweig, M. C. *et al. Science* **236**, 816-819 (1987).
13. Reth, M. G., Knieps, E., Wiese, P., Lobel, L. & Alt, F. W. *EMBO J.* (in the press).
14. Yuan, D. & Vitetta, E. S. *J. Immun.* **120**, 353-356 (1978).
15. Yuan, D., Gilian, A. C. & Tucker, P. W. *Fen. Proc.* **44**, 2652-2659 (1985).
16. Oi, V. T., Jones, P. P., Goding, J. W. & Herzenberg, L. A. *Curr. Top. microbiol. Immun.* **81**, 115 (1978).
17. Dialynas, D. P. *et al. J. Immun.* **131**, 2445 (1983).
18. Yelton, D. E., Desaymard, C. & Scharff, M. D. *Hybridoma* **1**, 5 (1981).
19. Kincaid, P. W., Lee, G., Sun, L. & Watanabe, T. *J. immunol. Meth.* **42**, 17-26 (1981).
20. Ledbetter, J. A. & Herzenberg, L. A. *Immunol. Rev.* **47**, 63 (1979).
21. Mulligan, R. C. & Berg, P. *Proc. natn. Acad. Sci. U.S.A.* **78**, 2072-2076 (1981).
22. Maki, R. *et al. Cell* **24**, 353-365 (1981).
23. Alt, F. *et al. EMBO J.* **3**, 1209-1219 (1984).

Apparent role of the macrophage growth factor, CSF-1, in placental development

Jeffery W. Pollard*§, Anna Bartocci*, Robert Arcucci†, Amos Orlofsky*, M. B. Ladner‡ & E. Richard Stanley*¶

* Department of Developmental Biology and Cancer, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York, New York 10461, USA

† Division of Pediatric Hematology and Oncology, Dana Farber Cancer Institute, The Children's Hospital, 44 Binney Street, Boston, Massachusetts 02115, USA

‡ Department of Molecular Biology, Cetus Corporation, 1400 53rd Street, Emeryville, California 94608, USA

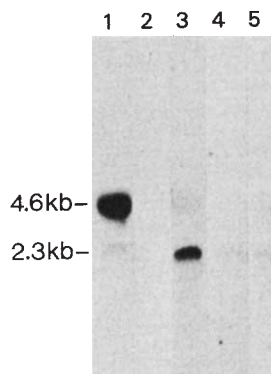
Colony stimulating factor-1 (CSF-1) is a glycoprotein growth factor required for the proliferation and differentiation of mononuclear phagocytic cells (reviewed in ref. 1). A 10,000-fold elevation of mouse uterine CSF-1 during pregnancy, suggested by studies of the bone marrow colony stimulating activity of uterine extracts^{2,3}, was recently demonstrated by radioimmunoassay (RIA)⁴. This increase and the observations that placenta and choriocarcinoma cell lines express *c-fms* messenger RNA and the *c-fms* proto oncogene product (CSF-1 receptor⁵) respectively^{6,7}, suggest an additional role for CSF-1 in pregnancy. We now show that uterine CSF-1 concentration is regulated by the synergistic action of female sex steroids, oestradiol-17 β (E₂) and progesterone (P) and that the elevation in CSF-1 concentration can be attributed to the preferential expression of an alternatively spliced CSF-1 mRNA by uterine glandular epithelial cells. These findings indicate that CSF-1, under hormonal influence, plays a role in placental development and function and that steroid hormones may regulate developmental processes via their effects on the expression of tissue-specific growth factors.

§ Present address: Department of Biochemistry, King's College, University of London, Campden Hill, London W87AH, UK.

¶ To whom correspondence should be addressed.

Fig. 1 Identification of CSF-1 mRNA by Northern blotting.

Methods. Total L cell or C3H/HeJ mouse RNA was extracted⁸ and 40 µg subjected to Northern blotting^{9,10}. Blots were probed with the insert of a 2.3-kb cDNA clone to murine CSF-1 mRNA (pGEM2mCSF-53), containing the entire coding sequence and labelled with ³²P-dCTP by random oligonucleotide priming¹¹. Hybridization was performed at 42 °C as described¹⁰ and blots were washed to a final stringency of 0.1 × SSC at 62 °C, prior to autoradiography. RNA was isolated from lane 1, L929 cells (50 h); lane 2, uterus (14 days); lane 3, uterus at day 19 of gestation (50 h); lane 4, placenta (14 days) and lane 5, uterus from P and E₂ (25 ng) treated mice that received a decidual stimulus (14 days) (figures in parentheses are autoradiographic exposure times). All experiments were repeated at least twice with comparable results.



To show that the increase in uterine CSF-1 in pregnancy was due to local synthesis, rather than to accumulation from the circulation, the uteri of pregnant and non-pregnant mice were examined for the presence of CSF-1 mRNA by Northern blot analysis (Fig. 1). Pregnant uteri contained a major (91%) 2.3-kb mRNA and a minor (9%) 4.6-kb species. In contrast, CSF-1 mRNA was not detected in the uteri of non-pregnant mice, even following sevenfold longer autoradiographic exposure times. While CSF-1 mRNA could be detected in placental samples at the longer exposure time, this was probably due to contaminating endometrial tissue that adhered to placentae during their dissection. These data indicate that CSF-1 is synthesized in pregnant uteri and in very low amounts, if at all, in non-pregnant uterus and placenta.

CSF-1 purified from pregnant mouse uterus and L cell conditioned medium, exhibited similar specific biological activities and molecular weights (data not shown). But, the proportion of the 2.3-kb and 4.6-kb mRNA species in pregnant uterus was the inverse of their proportion in mouse L cells (5% and 95%, respectively, Fig. 1). There is only one CSF-1 gene^{12,13}. The two major mouse mRNA species (2.3 kb and 4.6 kb) transcribed from this gene differ in their 3'-untranslated sequence^{13,14}. As CSF-1 mRNA from pregnant uterus is predominantly the 2.3-kb species whereas CSF-1 mRNA in fibroblasts is primarily the 4.6-kb form, the preferential expression of the smaller mRNA is probably due to cell-type specific regulation of mRNA splicing. Also of interest is the fact that the complementary DNA sequences indicate that the 4.6-kb, but not the 2.3-kb, species contains a consensus sequence (AUUUA) in the 3'-untranslated region¹⁴ that confers a short mRNA half life¹⁵. Thus it is possible that the mRNA synthesized by the uterus is long-lived and preferred as a requirement for sustained CSF-1 production during pregnancy.

To identify the uterine cells responsible for CSF-1 production, longitudinal uterine sections of 11-day pregnant mice were subjected to *in situ* hybridization with the CSF-1 complementary RNA probe. Corresponding bright and dark field photomicrographs (Fig. 2a, b) demonstrate the preferential localization of CSF-1 antisense RNA to the luminal and glandular secretory epithelium of the endometrium. Figure 2c and d show an absence of specific localization of autoradiographic grains in sections hybridized to the sense-strand of CSF-1 RNA. In higher magnification photomicrographs, it can also be seen that with the antisense but not sense RNA, the autoradiographic grains are preferentially localized to the columnar cells of the glandular epithelium (Fig. 2e, f). Also, *in situ* hybridization to sections of nonpregnant uterus revealed no significant signal over the luminal secretory epithelium, consistent with the Northern blot hybridiz-

ation analysis (Fig. 1) (data not shown). The results of this *in situ* hybridization analysis confirm that CSF-1 mRNA is induced and accumulates in the gravid uterus. They also demonstrate that CSF-1 mRNA preferentially accumulates in the endometrial secretory epithelium.

Human chorionic gonadotrophin (HCG) treatment, mimics the early increase in uterine CSF-1 observed in pregnancy⁴. As this effect was ablated by ovariectomy, and HCG is known to regulate ovarian production of sex steroid hormones, we investigated the effects of E₂ and P on uterine CSF-1 concentration in ovariectomized mice, using a protocol which permits implantation and decidualization to occur¹⁶. Either E₂ or P alone produced a small but significant increase in uterine CSF-1 concentration (Table 1). Increased uterine colony stimulating activity has been previously reported in mice receiving 100-fold higher concentrations of E₂ (ref. 18). In contrast, greater than additive effects were observed with a combination of P and E₂ (25 ng) (Table 1) or P and E₂ (50 ng) (data not shown), both of which caused a fourfold increase in uterine CSF-1 concentration. A 15-fold increase was observed in mice given a combination of P and E₂ and a decidual stimulus of arachis oil¹⁹ at day 3. No increase was observed in control mice given arachis oil alone (data not shown). The 10-ng E₂ dose was optimal for decidualization and only an occasional decidualization chamber was observed at 25 ng E₂ (data not shown). Thus CSF-1 production is not associated with the decidual cell production *per se* despite the fact that a decidual stimulus potentiates the increase in uterine CSF-1. Removal of E₂ from the protocol at day 4 reduced the uterine CSF-1 concentration to control levels or below (data not shown), indicating that the continual presence of E₂ is necessary for the response. In physiological terms, mice treated with a combination of P and E₂ on the day of killing, are equivalent to 9-day pregnant animals¹⁹ and their uterine CSF-1 concentration is only twofold lower than the concentration in 9-day pregnant uteri⁴. Serum CSF-1 concentrations in mice treated with both P and E₂ were not significantly different from those of carrier treated mice ($P > 0.01$, data not shown), consistent with a primary effect of these hormones on the uterine synthesis of CSF-1.

Direct evidence that HCG stimulates uterine CSF-1 production via progesterone was obtained by examining the effects of the antiprogesterin steroid RU 486 (ref. 20) (Table 2). The fourfold

Table 1 Effect of sex steroid hormones on uterine CSF-1 concentration

Dose of E ₂ (ng)	Hormone treatment	Uterine CSF-1 (fmol mg ⁻¹ tissue)
0	C	0.64 ± 0.01
0	P	0.83 ± 0.09
25	E ₂	0.88 ± 0.09
10	PE ₂	1.26 ± 0.14
10	†PE ₂	1.50 ± 0.29
10	*PE ₂	1.91 ± 0.55
25	PE ₂	2.67 ± 0.23
25	†PE ₂	3.14 ± 0.76
25	*PE ₂	9.33 ± 1.50

20–25 g 10–12 week old C3H/HeJ female mice were ovariectomized, rested for two weeks and primed with two subcutaneous injections of 100 ng E₂ (ref. 17). Three days after priming the combinations of P (500 µg), E₂ or arachis oil carrier (C) shown, were subcutaneously injected daily for 6 days. Mice marked * received a decidual stimulus of 10 µl of arachis oil intraluminally on day 3 of treatment. Control mice for the decidual response (†), received 10 µl of saline intraluminally on day 3 of treatment. One day after the last hormone injection, mice were decapitated, bled and their uteri processed for determination of CSF-1 using an RIA that only detects biologically active CSF-1 (ref. 4). Data shown are the means ± s.d. (≥ four mice per group) for a representative experiment. All experiments were repeated at least once with comparable results. The carrier treated group was significantly different from all experimental groups ($P < 0.05$, Student's *t*-test).

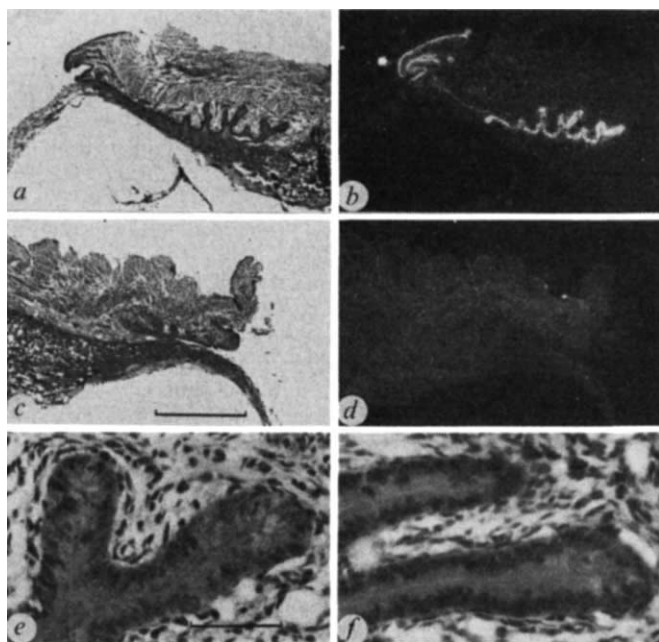


Fig. 2 Localization of CSF-1 RNA in tissue sections of gravid uterus by *in situ* hybridization. *a, c*, Brightfield photomicrographs showing convoluted endometrial glandular structures lying directly above the decidual layer of placenta as well as regions of endometrial secretory epithelium at the left and right borders of the tissue sections respectively. *b, d*, The same sections by darkfield photomicrography to demonstrate the preferential localization of CSF-1 RNA by the antisense RNA probe (*b*) and the low non-specific binding of the control sense probe (*d*). Bar equals 0.5 mm. *e, f*, Higher magnification photomicrographs of uterine glandular structures and surrounding tissue hybridized to the CSF-1 antisense (*e*) and sense (*f*) cRNAs respectively. 1,099 and 76 grains, respectively, are localized to the columnar epithelial cells in *e* and *f*. Bar = 0.1 mm. Exposure time, 136 h.

Methods. Uterus at day 11 of gestation was fixed, paraffin embedded, sectioned and hybridized to 32 P-labelled antisense or control sense-strand CSF-1 RNA (Arcaci, R. J., Croop, J., Horwitz, S. & Housman, D., manuscript submitted). Both probes were identically labelled and alkaline hydrolysed to a mean length of 150 base pairs. Probe specific activity and concentration were approximately 5×10^8 c.p.m. per μ g and about 0.15 μ g per ml of hybridization solution respectively.

increase in uterine CSF-1 in mice given HCG was almost completely abrogated by RU 486 administration, indicating an absolute requirement for P. To demonstrate that the increase in CSF-1, induced by treatment with P and E_2 , like that observed in pregnancy is due to increased CSF-1 synthesis, RNA extracted from the uteri of hormone-treated mice was examined by Northern blotting. Although the levels of induced CSF-1 mRNA were low compared with the levels in 16–21-day pregnant mice (Fig. 1), this was not unexpected as the CSF-1 concentration in the uteri of hormone-treated mice is ~ 100 -fold lower than in the uteri of mice in late pregnancy⁴. Thus, inhibition of the HCG effect on uterine CSF-1 by the antiprogesterin RU 486, coupled with the synergism between P and E_2 in the stimulation of uterine CSF-1 synthesis, suggest that CG (or luteinizing hormone) is involved (possibly with prolactin²²) in regulating the increase in uterine CSF-1 in pregnant mice via its stimulation of ovarian synthesis of P and E_2 .

At implantation, the embryonic tissue (trophoblasts) invades the endometrium and, together with the decidual cells of the uterus, develops into the extra-embryonic tissue, including the placenta. *In situ* hybridization studies indicate that *c-fms* mRNA is associated with trophoblasts in human²³ and mouse (data not shown) placentae. As CSF-1 mRNA was not detected in placenta by *in situ* hybridization it would appear that the CSF-1 pre-

Table 2 Inhibition of the HCG-stimulated increase in uterine CSF-1 concentration by the antiprogesterin, RU 486

Treatment	Uterine CSF-1 (fmol mg ⁻¹)
Saline	0.29 $\pm$ 0.03
HCG	*1.19 $\pm$ 0.07
HCG + RU 486	†0.42 $\pm$ 0.03
Saline + RU 486	0.32 $\pm$ 0.03

Adult C3H/HeJ female mice were intraperitoneally injected at daily intervals with HCG (25 IU per mouse; 12,000 IU per mg protein²¹) before being killed on day 5. RU 486 in arachis oil was subcutaneously injected (500 μ g per mouse) at daily intervals. CSF-1 concentrations (mean $\pm$ s.d. five mice per group) were determined as described in Table 1.

* Significantly different from saline-treated mice.

† Significantly different from the HCG treated group ($P < 0.001$, Student's *t*-test).

viously detected in the placenta (at $\leq 1/10$ of the uterine concentration in U per mg tissue)⁴ is not synthesized there but derives from the uterus. The localization of the site of CSF-1 synthesis to uterine luminal and glandular secretory epithelial cells is relevant to its putative role^{4,7} in the formation and maintenance of the placenta. First, the mitotic index and secretory activity of these cells are regulated by P and E_2 (refs 16, 24, 25). Second, populations of fetally-derived trophoblasts are stimulated to enter DNA synthesis by CSF-1 (ref. 26). Third, the close apposition of the uterine secretory epithelium to placental trophoblasts would permit their stimulation by released CSF-1 and even raises the possibility that CSF-1 presented at the epithelial cell surface in the form of the transmembrane CSF-1 precursor^{27,28} provides sites of attachment for the CSF-1 receptor-bearing trophoblasts in the formation of the placenta. These data, therefore, imply that CSF-1 synthesized in the uterus in response to P and E_2 regulates placental trophoblast proliferation and differentiation.

We thank Dr D. N. Ward for purified HCG and Roussel UCLAF Paris, France for their gift of RU 486. This work was supported by grants from the NIH, NATO, the UK Medical Research Council and the Charles H. Hood Foundation.

Received 13 July; accepted 19 October 1987.

- Stanley, E. R., Guilbert, L. J., Tushinski, R. J. & Bartelmez, S. H. *J. cell. Biochem.* **21**, 151–159 (1983).
- Bradley, T. R., Stanley, E. R. & Sumner, M. A. *Aust. J. exp. biol. med. Sci.* **49**, 595–603 (1971).
- Rosendaal, M. *J. cell. Sci.* **19**, 411–423 (1975).
- Bartocci, A., Pollard, J. W. & Stanley, E. R. *J. exp. Med.* **164**, 956–961 (1986).
- Sherr, C. J. *et al. Cell* **41**, 665–675 (1985).
- Muller, R. *et al. Molec. cell. Biol.* **3**, 1062–1069 (1983).
- Rettenmier, C. W. *et al. J. clin. Invest.* **77**, 1740–1746 (1986).
- Chirgwin, J. M., Przygla, R. J., MacDonald, R. J. & Rutter, W. J. *Biochemistry* **18**, 5294–5298 (1979).
- Thurston, C. T., Perry, C. R. & Pollard, J. W. in *Methods in Molecular Biology* Vol. 4, 1–13 (Humana, New Jersey, 1988).
- Pollard, J. W., Perry, C. R. & Thurston, C. F. in *Methods in Molecular Biology* Vol. 4, 14–29 (Humana, New Jersey, 1988).
- Fineberg, A. P. & Vogelstein, B. *Analyt. Biochem.* **132**, 6–13 (1983).
- Ladner, M. B. *et al. EMBO J.* **6**, 2693–2698 (1987).
- Rajavashisth, T. B. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 1157–1161 (1987).
- Ladner, M. B. *et al. Proc. natn. Acad. Sci. U.S.A.* (in the press).
- Shaw, G. & Kamen, R. *Cell* **46**, 659–667 (1986).
- Finn, C. A. & Martin, L. *J. Endocr.* **45**, 57–65 (1969).
- Cheng, S. V. Y., MacDonald, B. S., Clark, B. F. & Pollard, J. W. *Expl. cell. Res.* **160**, 459–470 (1985).
- Rosendaal, M. *Expl. Haemat.* **5**, 79–86 (1977).
- Finn, C. A. & Martin, L. *Biol. Reprod.* **7**, 82–86 (1972).
- Baulieu, E. E. in *The Antiprogesterin Steroid RU 486 and Human Fertility Control* (eds Baulieu, E. E. & Segal, S. J.) 1–25 (Plenum, New York and London, 1985).
- Canfield, R. E., Morgan, F. J., Kammerman, S., Bell, J. J. & Agosto, G. M. *Recent Prog. Horm. Res.* **27**, 121–164 (1971).
- Morishige, W. K. & Rothchild, I. *Endocrinology* **95**, 260–274 (1974).
- Hoshina, M., Hishio, A., Bo, M., Boime, I. & Mochizuki, M. *Acta obstet. gynec. Jpn* **37**, 2791–2798 (1985).
- Tachi, C., Tachi, S. & Lindner, H. R. in *Implantation of the Ovum* (eds Yoshinaga, K., Meyer, R. K. & Greep, R. O.) 43–80 (Harvard University Press, Cambridge, 1976).
- Zhinkin, L. N. & Samoskina, N. A. *J. embryol. exp. Morph.* **17**, 593–605 (1967).
- Athanasakis, I. *et al. J. Immun.* **138**, 37–44 (1987).
- Kawasaki, E. S. *et al. Science* **230**, 291–296 (1985).
- Rettenmier, C. W. *et al. Molec. cell. Biol.* **7**, 2378–2387 (1987).

HIV infection of primate lymphocytes and conservation of the CD4 receptor

Myra O. McClure*, Quentin J. Sattentau†, Peter C. L. Beverley‡, John P. Hearn§, Anthony K. Fitzgerald§, Arie J. Zuckerman|| & Robin A. Weiss*

* Chester Beatty Laboratories, Institute of Cancer Research, London SW3 6JB, UK

† Department of Genito-Urinary Medicine, Middlesex Hospital Medical School, London W1P 7LD, UK

‡ ICRF Human Tumour Immunology Group, Faculty of Clinical Sciences, University College School of Medicine, London WC1E 6JJ, UK

§ MRC/AFRC Comparative Physiology Group, Institute of Zoology, London NW1 4RY, UK

|| Department of Medical Microbiology, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK

The CD4 T-lymphocyte differentiation antigen is an essential component of the cell surface receptor for human immunodeficiency viruses (HIVs) causing AIDS (acquired immune deficiency syndrome) (refs 1–3). Peripheral blood lymphocytes of apes, New World and Old World monkeys express cell surface antigens homologous to CD4 of human T-helper lymphocytes. The cells of several of these species can be infected in short term culture with diverse strains of the type-1 or type-2 human immunodeficiency viruses (HIV-1 and HIV-2). HIV-1 is the prototype AIDS virus, and HIV-2 is the second type of AIDS virus, prevalent in West Africa⁴. Infection of the primate cells correlates with evolutionary conservation on CD4 of one particular epitope cluster, and is

inhibited by treatment of the cells with monoclonal antibodies to this epitope. The capacity of HIV to replicate in simian cells may provide a means for evaluating antiviral drugs and vaccines.

Using monoclonal antibodies (mAbs) to CD4, expression of the CD4 T-lymphocyte differentiation antigen can be identified by immunofluorescence and flow cytometric analysis. We have previously conducted epitope mapping of CD4 on T-cell leukaemia lines and on human peripheral blood lymphocytes (PBL) in relation to HIV binding, using a panel of 25 mAbs (ref. 5). In the present study, PBL from a wide range of primates (Table 1) were separated from aliquots of fresh heparinized blood taken from diagnostic samples to determine which species are susceptible to infection by HIV-1 and by HIV-2. The PBL of 14 of the non-human primate species were also studied to determine which epitopes of the CD4 molecule were conserved, and to correlate CD4 expression with HIV infection. Some mAbs against CD4 block the interaction of HIV-1 and HIV-2 with the human cell receptor^{5,6}. We therefore ascertained whether the same CD4 epitopes were essential for virus recognition in other primate species, and whether mAbs raised against them would block HIV infection.

PBL from a diverse range of primate species were exposed in culture to HIV-1 and, in some cases, to HIV-2 (Table 1). Cells of many species were susceptible to infection by and replication of HIV-1, as assayed by three independent methods: (1) syncytium induction in CD4-bearing indicator cells; (2) the appearance in the PBL of the gag antigen, p18, by immune fluorescence microscopy; and (3) the detection of reverse transcriptase activity in the culture supernatant taken one and two weeks following infection. All PBL were positive for HIV-1 infection by syncytial induction, although in some cases (PBL cultures of cynomolgus, golden lion marmoset, lemur and bush baby), very few syncytia were apparent, and virus replication

Table 1 Infection of primate PBL by HIV-1 and HIV-2

Group	Species	Common name	Syncytial induction*		p18 ⁺ HIV-1	Reverse transcriptase‡	
			HIV-1	HIV-2		HIV-1	HIV-2
Human	<i>Homo sapiens</i>	Man	++	++	++	600,100	122,683
Apes	<i>Pongo pygmaeus</i>	Orang-utan	++		++	39,856	
	<i>Gorilla gorilla</i>	Lowland gorilla	++	++	+	152,640	50,353
	<i>Pan troglodytes</i>	Chimpanzee	++	++	++	176,101	51,250
	<i>Macaca mulatta</i>	Rhesus	++	++	+	378,455	108,554
Old world monkeys	<i>Macaca fascicularis</i>	Cynomolgus	±		–	–164	
	<i>Papio papio</i>	Baboon	+	+	–	13,527	11,874
	<i>Callithrix jacchus</i>	Common marmoset	++	–	+	35,615	735
New world monkeys	<i>Leontideus rosalia</i>	Golden lion marmoset	±	++	–	675	734
	<i>Saguinus labiatus</i>	Red-bellied tamarin	++	–	+	17,996	
	<i>Saguinus imperator</i>	Emperor tamarin	++		±	599	545
	<i>Saguinus oedipus</i>	Cotton-top tamarin	++	–	±	19,872	4,819
	<i>Cebus capucinus</i>	Capuchin monkey	++	±	+	249,972	5,881
	<i>Aotus trivirgatus</i>	Owl monkey	++	+	+	182,161	104,748
	<i>Chiropotes albinus</i>	White-nosed saki	±		+	4,157	431
	<i>Galago crassicaudatus</i>	Thick-tailed bush baby	±		–	140	
	<i>Lemur macaco fulvus</i>	Brown lemur	±		–	92	
Rodent	<i>Mus musculus</i>	Mouse	–	–	–	–100	150

The primate cells used in this study are classified according to Napier and Napier²⁶. PBL were derived from aliquots of fresh heparinized 5 ml blood samples obtained for diagnostic purposes and prepared by density gradient centrifugation (2,000g for 30 min at 4°C) on Ficoll-Paque (Pharmacia, Sweden). Cells, initially stimulated with phytohaemagglutinin (1 µg ml⁻¹), were maintained in RPMI-1640 medium containing fetal calf serum (FCS, 20%) and recombinant interleukin-2 (IL-2, 50–100 U ml⁻¹ from Boehringer Mannheim). The HIV-1 strain used was HTLV-IIIRF isolated from a Haitian²⁷ and HIV-2 was the LAV-2 ROD strain⁴. PBL were infected separately with 10⁶ infectious units in the C8166 human cell line²⁸. Following infection for 1 h at 37°C, PBL were washed three times (1,500g, 5 min) to remove free virus and maintained in culture in the presence of IL-2. Samples of infected cells or supernatants from these cultures were assayed for viral replication by three methods:

* Induction of cytopathic effect by co-cultivation 7 d following infection with an equal volume and number of a CD4⁺ C8166 indicator cells⁷. Cultures were subsequently examined for multi-nucleated giant cell formation (syncytia).

† Indirect immunofluorescence microscopy by monoclonal antibody²⁹ for the detection of p18 in infected PBL was carried out on cells fixed in paraformaldehyde and permeabilized with Triton-X100 (ref. 30).

‡ Viral reverse transcriptase activity was assayed³¹ using poly(rA).oligo(dT)_{12–18} as a template-primer and poly(dA).oligo(dT)_{12–18} as a cellular DNA background control. The results were means of duplicate experiments and are shown in c.p.m. minus the poly(dA) background (usually <1,000 c.p.m.). Cell supernatants were monitored for reverse transcriptase activity at 1 and 2 weeks following infection. The time point with the highest activity is shown.

Table 2 Epitope expression of the CD4 antigen by FACS analysis

Monoclonal antibody	Human	Gorilla	Chimpanzee	Rhesus	Mandrill	Baboon	Common marmoset	Golden lion marmoset	Red-bellied tamarin	Emperor tamarin	Cotton-top tamarin	Capuchin	Owl Monkey	White-nosed saki
2D1	++	+	++	-	-	-	-	·	-	·	-	-	-	·
Leu3a	++	++	++	++	++	++	++	+	+	+	++	++	++	+
F101-5	++	++	++	++	++	++	++	++	++	++	++	++	++	++
F101-69	++	++	++	++	+	++	+	++	++	++	++	++	++	++
MT310	++	++	++	++	+	++	++	++	++	+	++	++	++	++
91D6	++	++	++	·	+	++	++	+	-	+	·	·	++	+
94B1	++	++	·	·	++	++	++	+	+	+	·	·	++	+
T4/19THY	++	++	++	++	+	++	++	+	++	+	++	++	++	-
13B.8.2	++	++	++	+	+	++	+	++	+	++	+	+	+	++
EDU-2	++	++	++	+	-	++	+	++	+	+	+	+	++	+
66.1	++	++	++	++	+	++	++	-	+	+	+	+	+	-
NUTH-1	++	++	++	++	-	++	-	++	-	-	+	+	++	-
T418/t3	++	++	++	-	++	+	-	-	-	-	-	-	-	+
BL410T4	++	++	++	+	++	+	-	·	-	·	-	+	+	·
CLB-T41	++	+	·	·	-	++	-	+	-	-	·	·	-	·
B264/123	++	++	·	·	-	-	-	-	-	-	-	·	-	-
MT151	++	-	++	-	+	-	-	-	-	-	-	+	-	-
MT321	++	++	++	-	-	-	-	+	+	-	+	-	+	-
VIT4	++	++	++	-	+	-	-	-	-	-	+	-	+	-
G19-2	++	·	++	++	·	·	-	·	-	·	-	-	+	·
OKT4	++	++	++	++	++	-	-	++	+	+	+	-	-	-
OKT4a	++	++	++	·	++	++	-	++	-	-	·	·	-	++
OKT4b	++	·	·	·	·	·	-	·	-	·	·	·	-	·
OKT4c	++	++	·	·	-	+	-	+	-	-	·	·	-	-
OKT4d	++	+	·	·	-	++	-	-	-	-	·	·	-	-
OKT4e	++	-	·	·	-	++	-	·	-	·	·	·	-	·
Poly CD4	++	++	++	+	++	++	++	++	+	+	+	+	++	++

50 μ l cell suspension (2.5×10^5 cells) was incubated with 50 μ l mAb at 5 μ g ml⁻¹, or 1:100 dilution of ascitic fluid for 30 min at 4 °C. The cells were washed 3 times in RPMI 1640 with 2% FCS and 0.1% sodium azide, and resuspended in 50 μ l goat anti-mouse immunoglobulin conjugated to fluorescein isothiocyanate (Nordic). After 30 min at 4 °C, the cells were washed as before and resuspended in phosphate-buffered saline (PBS) without Ca²⁺ or Mg²⁺ and containing 1% formaldehyde. The percentage and peak channel fluorescence were analysed on a FACS IV (Becton-Dickinson) 24 h after fixation of the cells. The degree of staining is represented as ++ strong reaction, + weak reaction, - no reaction, · not done. The polyvalent CD4 antiserum (polyCD4) was prepared in mice by hyperimmunization with BALB/3T3 mouse cells transfected and expressing the human CD4 gene¹² followed by adsorption of the sera with the untransfected line.

was not confirmed by the other two assays. This could reflect the higher level of sensitivity of the syncytial induction assay, although fusion of C8166 cells by exogenous, input virus rather than by replicating virus cannot be excluded. No syncytia were found in cultures where uninfected PBL had been incubated with CD4⁺ indicator cells, and these control cultures were also negative for p18 synthesis and reverse transcriptase activity. As a control, murine PBL treated in the same way were refractile to infection, which correlates with the lack of reactivity of any of the CD4 mAbs with mouse cells.

To confirm the results shown in Table 1, where cells were infected with a single isolate of HIV-1 or HIV-2, three strains of HIV-1 from the USA, Haiti and Zaire (HTLV-IIIB, HTLV-IIIRF and Z129) showing envelope divergence⁷, were used to infect PBL of chimpanzee, rhesus monkey, common marmoset, cotton-top tamarin and owl monkey. PBL of each primate species tested were readily susceptible to infection by each of these three diverse HIV-1 isolates. We also found that a second isolate of HIV-2 (CBL-20) infected the same range of PBL as LAV-2. For selected cells, taken from the gorilla, rhesus macaque and cotton-top tamarin, reverse transcriptase activity following infection with HIV-1 was monitored over the period of one month. Maximum activity at 1 or 2 weeks post-infection gradually declines over weeks 3 and 4, even when cultures were supplemented with fresh primate PBL.

Having demonstrated the susceptibility of many species of

primate PBL to infection by HIV-1 and HIV-2, it was of interest to characterize the epitopes of their CD4 antigen, which acts as the HIV-binding receptor on human cells^{1-3,5}. Table 2 shows that many primate species share a limited number of CD4 epitopes with man, and that generally, the greater the phylogenetic distance from man, the fewer epitopes are conserved. For example, 22 out of 24 mAbs tested (92%) recognized epitopes on gorilla PBL, 18 out of 24 (75%) on baboon cells and 10 out of 26 (38%) on common marmoset cells. Note that Leu3a and OKT4a, which recognize closely similar epitopes on CD4 of human cells⁵, can be distinguished by binding to PBL of several of the New World monkey species. Only one epitope cluster was consistently represented in all primate species tested; this was detected by the group of CD4 mAbs containing Leu3a, F101-5, F101-69, MT310 and 94B1.

To test whether HIV infection of non-human primate cells depends on CD4 expression, anti-CD4 mAbs were used to block infection, as previously described for human cells^{1,2}. We found (Table 3) that Leu3a mAb strongly inhibited infection by HIV-1 in each of the species tested, but the results with F101-5 and MT151 mAbs did not correlate precisely with epitope expression by fluorescence-activated cell sorting (FACS) analysis. Although F101-5 showed bright immunofluorescence on PBL of all species, it blocked HIV infection in only 3 of 5 species tested, implying that this epitope is not essential for HIV-CD4 interaction. MT151 belongs to a different epitope cluster which is less

Table 3 Inhibition of infection of primate cells by anti-CD4 mAbs

Species	Replication of HIV-1 in the presence of			
	Leu3a	F101-5	MT151	No mAb
Man	—	—	—	++
Owl monkey	—	—	+	++
Red-bellied tamarin	—	—	+	++
Golden lion marmoset	—	++	++	++
White-nosed saki	—	++	++	++

250 μ l suspension containing 10^5 PBL in RPMI 1640 with FCS (20%) and IL-2 (100 U) were incubated with 10^5 infectious units of HIV-1 in the presence of 20 μ g ml⁻¹ Leu3a, F101-5 or MT151, or without mAb. One week after infection the cells were washed and co-cultivated with an equal volume and number of C8166 cells. Syncytial formation was evaluated at 24 h. Where many syncytia were observed, as in the control without mAb, this was represented as ++, a significant reduction (<10) was scored as +, and — indicated total absence of syncytia.

well represented on New World monkey PBL. This mAb showed partial inhibition of two species (owl monkey and red-bellied tamarin) that were negative by FACS. This partial blocking of HIV-1 could arise from the high concentration of mAb used (20 μ g ml⁻¹, when 1 μ g ml⁻¹ is sufficient to block HIV-1 infection of human cells), as there may be a low-affinity interaction with CD4 not detected by immunofluorescence.

The mAbs used here have been categorized into two main groups, according to whether or not they blocked the binding of HIV-1 and HIV-2 to human T cells^{5,6}. In addition, by cross-competition assays using ¹²⁵I-labelled antibody, clusters of mAbs were defined which were either wholly distinct from one another, or which competed for the same epitope⁵. Using this panel of characterized mAbs, we have shown that only one CD4 epitope cluster was consistently present in all primate species tested; the mAbs in this cluster are in competition and are therefore directed against the same or overlapping epitopes. Furthermore, HIV-1, HIV-2 and the T-cell tropic simian immunodeficiency viruses SIV_{MAC}, SIV_{MAN} and SIV_{AGM} all use the CD4 receptor on human cells and are blocked by these mAbs (ref. 6), emphasizing the importance of this epitope cluster to HIV/SIV binding. Anti-idiotypic antibodies to Leu3a, a mAb recognizing this epitope, neutralize HIV-1 (refs 8, 9).

Phylogenesis has been considered in relation to T-antigen expression on PBL using anti-CD3, CD4 and CD8 mAbs (refs 10–12), and our study of CD4 extends this analysis. Our findings differ, however, from an earlier report¹¹ which used a panel of 9 anti-CD4 mAbs and found that the gorilla shared fewer epitopes with man than did two of the three macaques phenotyped (*Macaca mulatta* and *Macaca arctoides*). The conserved epitope cluster recognized by HIVs and SIVs is probably important in the natural function of CD4 in the immune system and other tissues.

Transfer and ectopic expression of the human CD4 gene in human cells renders those cells susceptible to HIV infection, whereas CD4 gene transfer to mouse cells confers sensitivity to HIV binding but not to infection¹³. We have shown that primate CD4 is a receptor for HIV, so the function of human CD4 expressed in cells more closely related to man than mouse is of interest. PBL from several primate species have been infected *in vitro*¹⁴ and *in vivo*^{15–18} with simian lentiviruses. SIV_{MAC} (formerly STLIV-III_{MAC}) isolated from sick rhesus macaques¹⁵, and SIV_{MAN} from the healthy sooty mangabey¹⁸, have properties similar to HIV and produce an AIDS-like disease in recipient macaques. *In vitro* studies show that SIV_{MAC} replicates efficiently in host macaque PBL and progressively less so in PBL of human, gibbon, ape and baboon, and not at all in PBL from the chimpanzee, the squirrel monkey or the cotton-top tamarin¹⁴. Also, some simian cells are permissive to replication of HIV-1 and transfer of the cloned proviral genome^{19–21}.

We have shown that a much wider host range exists than formerly supposed for the infection of non-human primate *in*

vitro PBL by HIV-1 and HIV-2, although human cells were more permissive to viral replication than monkey cells. Following HIV-1 inoculation *in vivo*, chimpanzees show seroconversion, transient depression of CD4 lymphocytes, and lymphadenopathy (in one animal), but have not developed AIDS to date^{22–24}. In a recent attempt to infect rhesus macaques with HIV-2, three out of eight animals seroconverted within 12 months, but have shown no symptoms of disease so far²⁵.

We are grateful to C. Mulder, J. N. Weber and C. Hetherington for advice and discussions, to R. S. Tedder for providing us with anti-p18 mAb, and to the following for virus isolates: R. C. Gallo (HTLV-III_B, HTLV-III_{RF}); D. Zagury (Z129) and L. Montagnier (LAV-2). The study was supported by the Wellcome Foundation, the Cancer Research Campaign and the Medical Research Council.

Received 17 August; accepted 23 October 1987.

1. Dalglish, A. G. *et al.* *Nature* **312**, 763–767 (1984).
2. Klatzmann, D. *et al.* *Nature* **312**, 767–768 (1984).
3. McDougal, J. S. *et al.* *Science* **231**, 382–385 (1986).
4. Clavel, F. *et al.* *Science* **233**, 343–346 (1986).
5. Sattentau, Q. L., Dalglish, A. G., Weiss, R. A. & Beverley, P. C. L. *Science* **234**, 1120–1123 (1986).
6. Sattentau, Q. J. *Int. Conf. AIDS III*, 160 (1987).
7. Weiss, R. A. *et al.* *Nature* **324**, 572–575 (1986).
8. Chan, T. C., Dreesman, G. R. & Kennedy, R. C. *Proc. natn. Acad. Sci. U.S.A.* **84**, 3891–3895 (1987).
9. Sattentau, Q. J., Weber, J. N., Weiss, R. A. & Beverley, P. C. L. *Int. Conf. AIDS III*, 160 (1987).
10. Letvin, N. L. *et al.* *Eur. J. Immun.* **13**, 345–347 (1983).
11. Jonker, M. & Nooij, F. J. M. in *Leucocyte Typing* (eds Bernard, A. *et al.*) 373–388 (Springer, Berlin, 1984).
12. Letvin, N. L. in *Leucocyte Typing* (eds Bernard, A. *et al.*) 389–393 (Springer, Berlin, 1984).
13. Maddon, P. *et al.* *Cell* **47**, 333–348 (1986).
14. Kannagi, M., Yetz, J. M. & Letvin, N. L. *Proc. natn. Acad. Sci. U.S.A.* **82**, 7053–7057 (1985).
15. Letvin, N. L. *et al.* *Science* **230**, 71–73 (1985).
16. Baskin, G. B. *et al.* *J. natn. Cancer Inst.* **77**, 127–139 (1986).
17. Fultz, P. N. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **83**, 5286–5290 (1986).
18. Murphy-Corb, M. *et al.* *Nature* **321**, 435–437 (1986).
19. Levy, J. A. *et al.* *Virology* **147**, 441–448 (1985).
20. Levy, J. A., Cheng-Mayer, C., Dina, D. & Luciw, P. A. *Science* **232**, 998–1001 (1986).
21. Adachi, A. *et al.* *J. Virol.* **59**, 284–291 (1986).
22. Alter, H. J. *et al.* *Science* **226**, 549–552 (1984).
23. Fultz, P. N. *et al.* *J. Virol.* **58**, 116–124 (1986).
24. Nara, P. L. *et al.* *J. Virol.* **61**, 3173–3180 (1987).
25. Fultz, P. N., Switzer, W. M. & Montagnier, L. *Int. Conf. AIDS III*, 212 (1987).
26. Napier, J. R. & Napier, P. H. *A Handbook of Living Primates* (Academic, London, New York, 1967).
27. Gallo, R. C. *Science* **224**, 500–503 (1984).
28. Clapham, P. R. *et al.* *Virology* **158**, 44–51 (1987).
29. Ferns, B., Tedder, R. S. & Weiss, R. A. *J. gen. Virol.* **68**, 1543–1551 (1987).
30. Matlin, K. S., Reggio, H., Helenius, A. & Simons, K. *J. Cell Biol.* **91**, 601–613 (1981).
31. Hoffman, A. D., Banapur, B. & Levy, J. A. *Virology* **147**, 326–335 (1985).

Anti-termination of transcription within the long terminal repeat of HIV-1 by *tat* gene product

Shaw-Yi Kao*, Andrew F. Calman*, Paul A. Luciw† & B. Matija Peterlin*‡

* Howard Hughes Medical Institute, Department of Medicine and Department of Microbiology and Immunology, University of California, San Francisco, California 94143, USA

† Department of Medical Pathology, University of California, Davis, California 95616, USA

Human immunodeficiency virus-1 (HIV-1) gene expression is controlled by cellular transcription factors and by virally encoded *trans*-activation proteins of the HIV-1 *tat* and *art*/*tr*s genes, which are essential for viral replication^{1,9–11}. *Tat* *trans*-activates HIV-1 gene expression by interacting with the *trans*-acting response element (TAR) located within the HIV-1 long terminal repeat (LTR) (ref. 2). In transient expression assays, *tat* mediates its effects largely by increasing the steady-state levels of messenger RNA species that contain the TAR sequence at or near their 5' ends^{2–4}, suggesting a function for *tat* either in transcription or in subsequent

‡ To whom correspondence should be addressed.

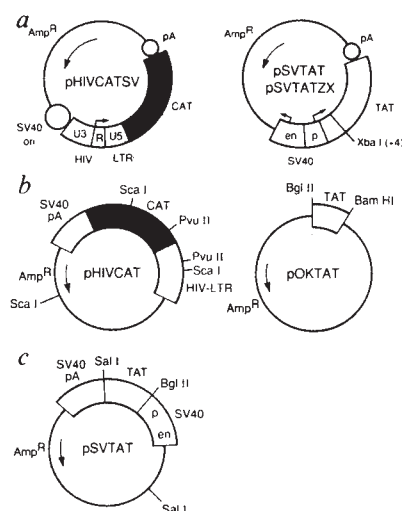


Fig. 1 Plasmids and their restriction sites. *a*, Plasmids used in transient transfection assays. pHIVCATSV contains the HIV-1 LTR linked to the CAT gene. pSVTAT encodes functional *tat*. pSVTATZX, derived from pSVTAT, contains a frameshift mutation which inactivates *tat*. All three plasmids contain the SV40 origin of replication, which permits their replication in COS7 cells. *b*, Plasmids used for Southern blot analysis. pHIVCAT is similar to pHIVCATSV but lacks the SV40 origin of replication. pOKTAT contains the *tat* gene. Other plasmids (not shown): pHIVCATSV2 is similar to pHIVCATSV, except that the SV40 origin of replication is downstream of the CAT gene. pSVΔTAT, derived from pSVTAT, contains a deletion which inactivates *tat*. Key: en, transcriptional enhancer. p, transcriptional promoter. pA, polyadenylation site. ori, origin of replication. Large arrows, bacterial ampicillin resistance gene; small arrows, sites and direction of initiation of transcription.

Methods. pHIVCATSV contains 4.5 kilobase (kb) pairs and was constructed by ligation of the 2,200-bp *EcoRV*-*Bam*HI fragment of pHIVCAT (previously designated³ TAR-1), which contains HIV-1 LTR, CAT and SV40 polyadenylation sequences, into the *Xba*I and *Bam*HI sites of pSVT-2 (ref. 25). pSVTAT (3.8 kb; previously designated TAT-1), which expresses the *tat* protein, has been described³. pSVTATZX was derived from pSVTAT by linearizing with *Xba*I (position +21 in *tat*), filling recessed ends, and religating. This introduced four additional nucleotides into the *tat* sequence, creating a frameshift which inactivates *tat*. pOKTAT is a pUC-based plasmid with a few additional restriction sites into which the *tat* gene was cloned. pHIVCATSV2 was derived from pHIVCAT by ligation of a 342-bp fragment containing the SV40 replication origin downstream of the CAT gene. pSVΔTAT was constructed from pSVTAT by deletion of the 83 bp *Xba*I-*Mlu*I fragment in the *tat* gene. All plasmids are based on the vector pML (ref. 26). Structure of all plasmids was verified by restriction mapping and/or DNA sequencing. Production of functional *tat* mRNA from pSVTAT, and of non-functional *tat* mRNA from pSVTATZX and pSVΔTAT, was verified by CAT enzyme assay and *tat* RNA analysis in cotransfected cells (Fig. 2 and data not shown).

RNA processing. The *tat* gene could also facilitate translation of mRNA containing the TAR sequence⁵⁻⁸. To determine the mechanism of *trans*-activation by *tat*, we analysed the structure and rate of synthesis of RNA species directed by the HIV-1 LTR in transient expression assays both in the presence and absence of *tat*. Although the rate of HIV-1 transcription initiation was not affected by *tat*, transcriptional elongation beyond position +59 was seen only in the presence of *tat*. Thus, *tat* *trans*-activates HIV-1 transcription by relieving a specific block to transcriptional elongation within the TAR sequence.

In human immunodeficiency viruses, viral gene expression and replication are dependent on virally encoded *trans*-activation proteins^{1,9-11}. Infectivity of HIV-1 is abolished by deletions within the gene encoding the viral *trans*-activator *tat*; when *tat* is provided in *trans*, viral gene expression is restored¹² by a

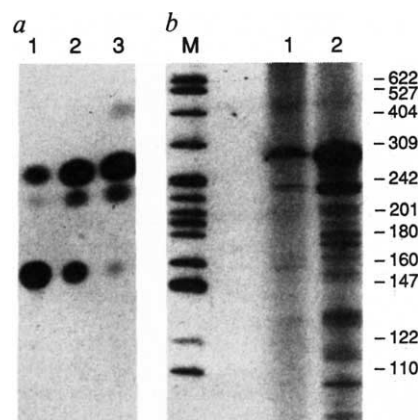


Fig. 2 CAT enzymatic assay and primer extension analysis of cotransfected COS7 cells. COS7 cells were cotransfected with pHIVCATSV and either pSVTATZX or pSVTAT. Cells were collected for CAT enzyme assays and primer extension experiments 48 h later. *a*, CAT enzyme assay of cellular lysates from COS7 cells cotransfected with pHIVCATSV and pSVTATZX (lane 1) or with pHIVCATSV and pSVTAT (lanes 2 and 3). Assays contained 5 μ g (lanes 1 and 3) or 1 μ g (lane 2) protein from cellular lysates. CAT converts ¹⁴C-chloramphenicol to its acetylated forms, which migrate faster on thin layer chromatography plates. *b*, primer extension assay. CAT primer (20 nucleotides, positions +50 to +70 in the CAT gene) was used to prime reverse transcription from 10 μ g total RNA from these cotransfected cells. Lane M, molecular weight markers (pBR322 digested with *Msp*I). Fragment sizes (in nucleotides) are indicated at the right. Lane 1, COS7 cells cotransfected with pHIVCATSV and pSVTATZX. Lane 2, COS7 cells cotransfected with pHIVCATSV and pSVTAT. The major band of 280 nucleotides corresponds to transcripts which initiate at the HIV-1 LTR cap site³.

Methods. COS7 cells were grown in 10-cm tissue culture dishes. Supercoiled plasmid DNA (10 μ g) was transfected in 5 ml Dulbecco's modified Eagle medium containing 400 μ g ml⁻¹ DEAE-dextran, 0.1 mM chloroquine diphosphate and 50 mM Tris-HCl, pH 7.5. After a 3 h incubation, cells were washed with serum-free medium and incubated for 48 h in medium containing 10% fetal bovine serum, then scraped from the dishes. A small fraction was lysed by three freeze-thaw cycles for the CAT enzyme assay. RNA was extracted from the remaining cells by the guanidinium isothiocyanate/CsCl method²⁷. CAT assays³ were performed using a 10 min incubation at 37 °C. After thin layer chromatography on silica plates, spots containing monoacetylated and non-acetylated ¹⁴C-chloramphenicol were excised and counted by liquid scintillation to determine CAT activity. Primer extensions were performed as described³, and quantified by scanning densitometry of the autoradiograph.

mechanism which is not fully understood. We studied the mechanism of HIV-1 *trans*-activation by *tat* in COS7 African green monkey kidney cells. These cells express the T antigen of SV40 and permit efficient replication and expression of plasmids which contain the SV40 origin of replication¹³. We have constructed several such plasmids (Fig. 1). pHIVCATSV contains the HIV-1 LTR linked to the bacterial chloramphenicol acetyltransferase (CAT) reporter gene. pSVTAT (ref. 3) directs the synthesis of functional *tat* protein, whereas pSVTATZX, which contains a frameshift mutation at position +21 of the *tat* gene, encodes a non-functional *tat* protein. pHIVCATSV was cotransfected with either pSVTAT or pSVTATZX into COS7 cells. Using a transient expression assay, we quantified the amounts of steady-state CAT protein and mRNA, measured rates of transcriptional initiation and elongation from the HIV-1 LTR, and mapped HIV-1 transcripts synthesized in the presence and absence of functional *tat*.

Figure 2 shows that when COS7 cells were cotransfected with pHIVCATSV and pSVTATZX, amounts of both CAT enzyme

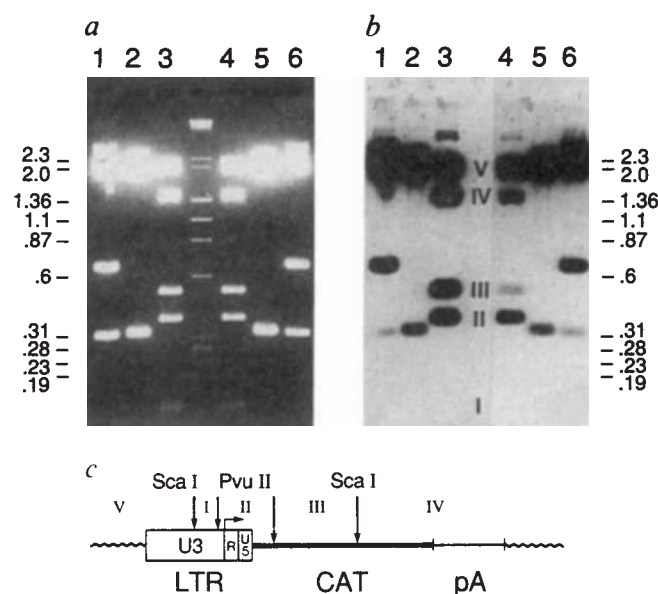


Fig. 3 Nuclear run-on transcription analysis of cotransfected COS7 cells. *a*, Ethidium bromide staining of DNA fragments in a 2% agarose gel used for Southern blotting. Lanes 1 and 6, pSVTAT digested with *Bgl*II and *Sal*I. Lanes 2 and 5, pOKTAT digested with *Bgl*II and *Bam*HI. Lanes 3 and 4, pHIVCAT digested with *Sca*I and *Pvu*II. DNA markers are from bacteriophage λ digested with *Hind*III and bacteriophage ϕ X174 digested with *Hae*III; lengths are given in kb. *b*, Hybridization of radiolabelled nascent transcripts to Southern blot of Fig. 3*a*. Lanes 1 to 3, COS7 cells cotransfected with pHIVCATSV and pSVTAT. Lanes 4 to 6, COS7 cells cotransfected with pHIVCATSV and pSVTATZX. Fragments labelled I-V represent restriction fragments of pHIVCAT, as shown in Fig. 3*c*. *c*, Restriction fragments of pHIVCAT digested with *Pvu*II and *Sca*I. Fragment I, 120 bp *Sca*I-*Pvu*II band from the U3 region; fragment II, 360 bp *Pvu*II-*Pvu*II band from the R, U5, and 5'CAT regions; fragment III, 520 bp *Pvu*II-*Sca*I band from the CAT region; fragment IV, 1,400 bp *Sca*I-*Sca*I band from the CAT, SV40 polyA, and pML regions; fragment V, 2,200 bp *Sca*I-*Sca*I band from the pML and U3 regions.

Methods. Restriction digests of 5 μ g plasmid DNA were electrophoresed on 2% agarose gels, denatured *in situ* and transferred to nylon membranes (Micron Separations) using 1M ammonium acetate. Because the pSVTAT digest was incomplete, less of the 300 bp *tat* fragment was generated in lanes 1 and 6 than in lanes 2 and 5. The DNA fragments were immobilized on the filters by UV cross-linking, which ensures the retention of small DNA fragments²⁸. COS7 cells were cotransfected as in Fig. 2. After a 48 h incubation, the cells were permeabilized with 200 μ g ml⁻¹ digitonin, and nuclei were incubated with α -³²P-UTP for 15 min at 25 °C. RNA was extracted with guanidinium HCl, and 2 $\times$ 10⁶ c.p.m. RNA was hybridized to each filter for 48 h. Filters were then washed, treated with 10 μ g ml⁻¹ RNase A for 10 min at 22 °C, and autoradiographed as described¹⁴. Results were quantified by scanning densitometry of several different exposures of each autoradiograph, normalizing to the *tat* signals. The data shown are representative of five independent experiments. Incorporation of α -³²P-UTP was linear from 0-45 min, and an identical pattern of hybridization was obtained using RNA synthesized in either a 5 min or a 15 min reaction.

activity and correctly initiated CAT mRNA were low. In COS7 cells cotransfected with pHIVCATSV and pSVTAT, the concentrations of CAT protein and RNA were increased 40-fold and 33-fold respectively. The extent of *trans*-activation of the HIV-1 LTR by *tat* in COS7 cells was equivalent when similar plasmid constructions which do not contain the SV40 origin of replication were used (data not shown). These results validate the use of replicating plasmids and COS7 cells for the study of HIV-1 *trans*-activation, and confirm and extend our previous analysis of HIV-1 *trans*-activation in human T cells, HeLa cells and Syrian hamster pancreatic cells³. Because increases in CAT

enzyme activity paralleled the increases in steady-state levels of mRNA transcribed from the HIV-1 LTR, we conclude that *tat* must act at a pretranslational step.

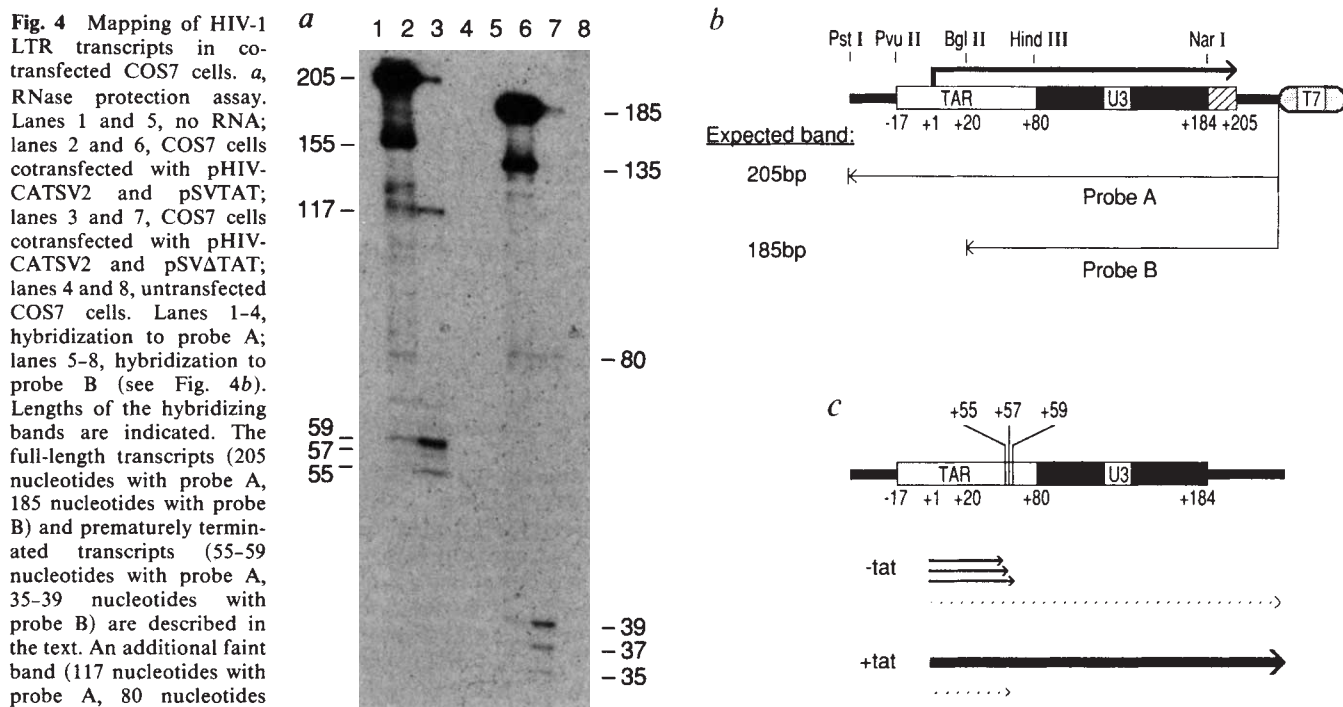
We used an *in vitro* nuclear run-on transcription assay¹⁴ to study the effect of *tat* on the rates of HIV-1 transcriptional initiation and elongation. Radiolabelled nascent transcripts from cotransfected COS7 cells were hybridized to Southern blots of immobilized *tat*, CAT, and HIV-1 LTR sequences (Fig. 1*b* and Fig. 3). Hybridization to the 300-base-pair (bp) *tat* band served as control for both transfection efficiency and total incorporation of radioactive label in each reaction (Fig. 3*b*, lanes 1 and 6).

To compare the transcription rates of different regions of the HIV-1 LTR, pHIVCAT (Fig. 1*b*) was digested with *Sca*I and *Pvu*II (Fig. 3*a*, lanes 3 and 4). These two restriction enzymes yield five DNA fragments representing sequences upstream and downstream from the site of initiation of HIV-1 transcription (Fig. 3*c*). No hybridization to the HIV-1 LTR U3 region (fragment I) was detected (Fig. 3*b*); thus, little or no transcription occurs from upstream of the cap site. RNA transcribed both in the presence and absence of *tat* hybridized strongly to fragment II, which contains sequences from the R and U5 regions of the HIV-1 1 LTR and the CAT gene (Fig. 3*b* and *c*). In contrast, hybridization to fragment III, which contains only CAT sequences, was 47-fold higher in cells cotransfected with pHIVCATSV and pSVTAT than in cells cotransfected with pHIVCATSV and pSVTATZX. In both samples, equivalent hybridization was observed to pML sequences transcribed from the SV40 late promoter contained in pSVTAT and pSVTATZX (fragments IV and V). We conclude that *tat* does not affect the rate of HIV-1 initiation of transcription, but is required for transcriptional elongation through the HIV-1 LTR. This suggests that only short transcripts are synthesized in the absence of *tat*.

To identify and map these predicted short transcripts, we used an RNase protection assay to analyse the HIV-1 LTR RNA species synthesized in the presence or absence of *tat* (Fig. 4*a*). ³²P-labelled HIV-1 LTR anti-sense RNA probes A and B (Fig. 4*b*) were hybridized to RNA from COS7 cells cotransfected with pHIVCATSV2 and either pSVTAT or pSVΔTAT (Fig. 1, legend), digested with RNases A and T1, and analysed by polyacrylamide gel electrophoresis. In the presence of *tat*, the predominant protected band (205 nucleotides with probe A, 185 nucleotides with probe B) corresponded to HIV-1 LTR transcripts which were initiated at position +1 and elongated beyond position +184 (Fig. 4*a*, lanes 2 and 6). An additional band (155 nucleotides with probe A, 135 nucleotides with probe B) of variable intensity, even for the same RNA samples, may have arisen from cleavage at an RNase-sensitive site created by the formation of a stable hairpin structure near the 5' end⁴.

In the absence of *tat*, the amount of the full-length HIV-1 LTR transcript was decreased 50-fold (Fig. 4*a*, lanes 3 and 7). We detected several shorter transcripts in these cells (59, 57 and 55 nucleotides with probe A; 39, 37 and 35 nucleotides with probe B) which initiate at position +1 and elongate to positions +55, +57 and +59, respectively. We compared the intensity of the 55-59 nucleotide bands to that of the 205 nucleotide band by densitometer scanning of the autoradiograph in Fig. 4*a*. Whereas in the absence of *tat* ~87% of the correctly initiated transcripts terminated at positions +55 to +59, in the presence of *tat* more than 99% of transcripts were full length. Interestingly, the most prominent transcription endpoint (position +59) corresponds to the junction of two stable RNA stem-loop structures⁴. These structures have a calculated free energy of -37 kcal mole⁻¹, and form under physiological conditions *in vitro*⁴. We conclude that *tat* is required for transcriptional elongation beyond position +59 of the HIV-1 TAR element.

Previously, we demonstrated that *trans*-activation by *tat* increases the steady-state levels of HIV-1 mRNA (ref. 3). Here, we extend these findings by showing that in the absence of *tat*, transcripts that initiate at the HIV-1 cap site end within TAR



Methods. COS7 cells were transfected as in Fig. 2. After 48 h, total cellular RNA was prepared by hot phenol extraction and ethanol precipitation³⁰. ³²P-labelled RNA probes were synthesized from plasmid pLG1 using α -³²P-UTP and bacteriophage T7 RNA polymerase and gel-purified as previously described³. Hybridization was carried out as described³, except that hybridizations contained 10⁵ c.p.m. of probe and 5 μ g total cellular RNA in a volume of 15 μ l. Following hybridization, 75 μ l RNase digestion buffer (10 μ g ml⁻¹ RNase A, 1,300 units ml⁻¹ RNase T1, 10 mM Tris-HCl, pH 7.5, 300 mM NaCl, 5 mM EDTA) were added and samples were incubated for 1 h at 22 °C. RNases were then inactivated by proteinase K digestion and the ethanol-precipitated RNA samples were analysed by electrophoresis on an 11% polyacrylamide sequencing gel. After autoradiography, the relative molar quantities of the various RNA species were quantified by scanning densitometry. Results were corrected for the length of each hybridizing species as ³²P incorporation is proportional to probe length. Identical results were obtained when pHIVCATSV and pSVTATZX were used in place of pHIVCATSV2 and pSVΔTAT, respectively.

between positions +55 and +59. Because these transcripts accumulate in the cytoplasm (M. J. Selby and B.M.P., data not shown), they cannot arise simply from pausing of RNA polymerase, but must result from termination of transcription. We cannot formally exclude the possibility that transcription continues through position +59 in the absence of *tat*, and that the short transcripts arise from cleavage of the primary transcript and subsequent degradation of the 3' sequences. But such degradation would have to be extremely rapid, because little or no hybridization to CAT sequences was observed with nascent transcripts synthesised in a 5 or 15 minute transcription reaction in the absence of *tat* (Fig. 3*b*). Thus, *tat* acts as a transcriptional anti-terminator. This conclusion is supported by the fact that HIV-1 TAR, like the target sequences of prokaryotic anti-terminators, functions only when positioned in its native orientation, downstream of the site of transcriptional initiation^{3,4,15,16}.

In a variety of cells the effect of *tat* on the amount of steady-state HIV-1 mRNA can account for *trans*-activation²⁻⁴. An effect of *tat* on translation of HIV-1 mRNA has also been proposed⁵⁻⁸. In COS7 cells, we observed no effect of *tat* upon translation of HIV-1 mRNA as the increases in CAT protein and mRNA levels were equivalent. The run-on transcription experiments suggested that the rate of transcription initiation from the HIV-1 LTR was

not affected by *tat*. In the RNase protection assay, however, the total molar quantity of correctly initiated transcripts was about five times lower in the absence of *tat* (by densitometry of gel shown in Fig. 4*a*). This apparent discrepancy could be due to instability of the short transcripts *in vivo*, or to their selective loss during RNA isolation. Also, the length and secondary structure of these RNA species may interfere with hybridization and RNase digestion, leading to underestimation of their abundance. Thus, in the absence of data which directly implicate *tat* in the control of other transcriptional or post-transcriptional events, we conclude that the main and perhaps only function of *tat* in *trans*-activation is to relieve a specific block to transcriptional elongation.

Transcriptional termination and anti-termination are important regulatory mechanisms in many prokaryotic¹⁵⁻¹⁹ and eukaryotic¹⁸⁻²² genes, including the proto-oncogenes *c-myc* and *c-myb* (refs 23, 24). The study of transcriptional anti-termination in HIV-1 has the advantage that both the *trans*-acting factor (*tat*) and the *cis*-acting sequence (TAR) are known. Analysis of the mechanism of *trans*-activation by *tat* could also be important to the development of drugs designed to block anti-termination by *tat*.

We thank Kathy Jones, Suzanne Mansour, Mark Selby,

Charles Steinberg, Harold Varmus, Matthias Wabl and Michael Walker for discussions and helpful suggestions, Karen Shaw for technical assistance and Michael Armanini for preparation of the manuscript. This work was supported in part by grants from the State of California University wide Task Force on AIDS to BMP and PAL.

Received 2 September; accepted 22 October 1987.

1. Chen, I. S. Y. *Cell* **47**, 1-2 (1986).
2. Rosen, C. A., Soderroski, J. G. & Haseltine, W. A. *Cell* **41**, 813-823 (1985).
3. Peterlin, B. M., Luciw, P. A., Barr, P. J. & Walker, M. D. *Proc. natn. Acad. Sci. U.S.A.* **83**, 9734-9738 (1986).
4. Muesing, M. A., Smith, D. H. & Capon, D. J. *Cell* **48**, 691-701 (1987).
5. Cullen, B. *Cell* **46**, 973-982 (1986).
6. Wright, C. M., Felber, B. K., Paskalis, H. & Pavlakis, G. N. *Science* **234**, 988-992 (1986).
7. Rosen, C. A. *et al. Nature* **319**, 555-559 (1986).
8. Feinberg, M. B., Jarrett, R. F., Aldovini, A., Gallo, R. C. & Wong-Staal, F. *Cell* **46**, 807-817 (1986).
9. Soderroski, J. G. *et al. Science* **229**, 171-173 (1985).
10. Arya, S., Guo, C., Josephs, S. & Wong-Staal, F. *Science* **229**, 69-73 (1985).

11. Soderroski, J. G., Patarca, R., Rosen, C. A., Wong-Staal, F. & Haseltine, W. A. *Science* **229**, 74-77 (1985).
12. Dayton, A., Soderroski, J. G., Rosen, C. A., Goh, W. C. & Haseltine, W. A. *Cell* **44**, 941-947 (1986).
13. Gluzman, Y. *Cell* **23**, 175-182 (1981).
14. Mansour, S. L., Grodzicker, T. & Tjian, R. *Molec. cell. Biol.* **6**, 2684-2694 (1986).
15. Barik, S., Ghosh, B., Whalen, W., Lazinski, D. & Das, A. *Cell* **50**, 885-899 (1987).
16. Grayhack, E. J., Yang, Y., Lau, L. F. & Roberts, J. W. *Cell* **42**, 259-269 (1985).
17. Yanofsky, C. *Nature* **289**, 751-758 (1981).
18. Von Hippel, P. H., Bear, D. H., Morgan, W. D. & McSwiggen, J. A. *Rev. Biochem.* **53**, 389-446 (1984).
19. Platt, T. A. *Rev. Biochem.* **55**, 339-372 (1986).
20. Dedrick, R. L., Kane, C. M. & Chamberlin, M. J. *J. biol. Chem.* **262**, 9098-9108 (1987).
21. Reines, D., Wells, D., Chamberlin, M. J. & Kane, C. M. *J. molec. Biol.* **196**, 299-312 (1987).
22. Caci, G., Guise, J. W., McDevitt, M. A., Tucker, P. W. & Nevins, J. R. *Genes Dev.* **1**, 471-481 (1987).
23. Bentley, D. L. & Groudine, M. *Nature* **321**, 702-706 (1986).
24. Bender, T. P., Thompson, C. B. & Kuehl, W. M. *Science* **237**, 1473-1476 (1987).
25. Rio, D. C. & Tjian, R. *Cell* **32**, 1227-1240 (1983).
26. Lusky, M. & Botchan, M. *Nature* **293**, 79-81 (1981).
27. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. *Biochemistry* **18**, 3294-3299 (1979).
28. Church, G. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1991-1995 (1984).
29. Sanchez-Pescador, R. *et al. Science* **227**, 484-492 (1985).
30. Queen, C. & Stafford, J. *J. molec. Biol.* **4**, 1042-1049 (1984).

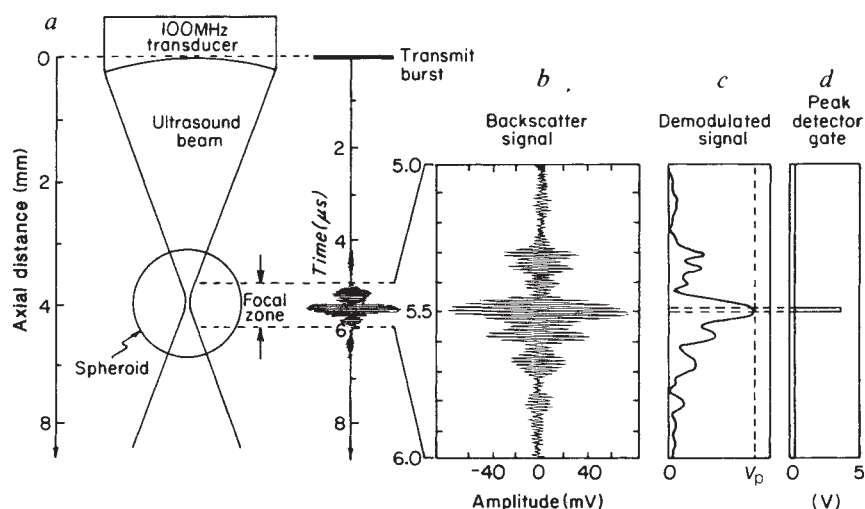
Ultrasound backscatter microscopy images the internal structure of living tumour spheroids

M. D. Sherar, M. B. Noss & F. S. Foster

Ontario Cancer Institute, 500 Sherbourne Street, Toronto, Ontario M4X 1K9, Canada

Ultrasound microscopes have the potential for imaging structure at depth in thick specimens, yet this is not possible in biological specimens using conventional ultrasound transmission or reflection methods¹⁻³. But, subsurfacing imaging is possible with ultrasound if a backscatter (pulse-echo) technique, similar to that used in medical imaging, is used⁴. The central problem of extending backscatter imaging to ultrasound microscopy has been the development of high frequency (>100 MHz) transducers with sufficient bandwidth and sensitivity to detect the low levels of backscatter from biological materials. We recently reported the development of such a transducer⁵ which we have now incorporated into a new ultrasound backscatter microscope capable of providing tomographic images at depths of up to 4 mm in biological specimens. Here we present the first ultrasound backscatter micrographs of living biological specimens. The benefits of this technique are demonstrated by its application to imaging the internal structures of living tumour spheroids showing striking contrast between the necrotic core and the viable rim of the spheroid.

Fig. 1 The ultrasound backscatter microscope and the method of image acquisition. A short, 130 V, 100 MHz electrical burst is transformed by the transducer into a short 100 MHz ultrasound pulse. The pulse is brought to a focus in the specimen (a) and the backscattered ultrasound (b) is detected at a time corresponding to the depth at which the scattering occurred. Note that the signal amplitude rises to a maximum at a time corresponding to scattering at the transducer focus, where the beam is most intense. This signal is demodulated (c) and peak detected in a time window defined by the gate signal (d). The peak backscatter signal, V_p , is transferred to a computer and is displayed as a brightness value in the image. A complete image is formed by scanning the ultrasound beam in two dimensions across the plane of interest providing 256×256 measurements, a process which takes approximately 10 min. Planes at different depths are imaged by moving the transducer such that the focus is in the plane of interest and repeating the above process.



We have applied a particular mode of ultrasound backscatter imaging called a C-scan⁶ to ultrasound microscopy. The microscope geometry and method of imaging are shown in Fig. 1. An ultrasound pulse, generated by the transducer, is brought to a focus in the specimen. The backscattered ultrasound, detected by the same transducer, is time-gated such that only scatter from the focal plane is accepted. The amplitude of this scatter signal is displayed as a brightness value in the image. A complete image is formed by scanning the specimen in two dimensions. The lateral resolution of the microscope is determined by the product of the average wavelength of the ultrasound pulse and the *f*-number (focal length/diameter) of the transducer. Our transducer is operated at 100 MHz with a wavelength of approximately 15 μm in biological specimens and has an *f*-number of 1.33 giving a theoretical resolution of 19.6 μm. The actual resolution (full width at half maximum) was measured to be 17.5 μm which is in good agreement with the theoretical prediction. The thickness of a tomographic slice is determined by the length of the ultrasound pulse. Our high bandwidth transducer is able to produce pulses which are short enough to give image slice thicknesses down to 28 μm. Further details, particularly on the transducer, are given elsewhere⁵.

We have applied the ultrasound backscatter microscope to imaging the internal structure of viable multicellular spheroids; a biological system of well known structure on which the microscope could be tested. Spheroids are clonal aggregates of tumour cells which serve as useful *in vitro* models of tumour micro-regions and early avascular tumour growth. As a spheroid grows, the central regions become depleted of oxygen and nutrients

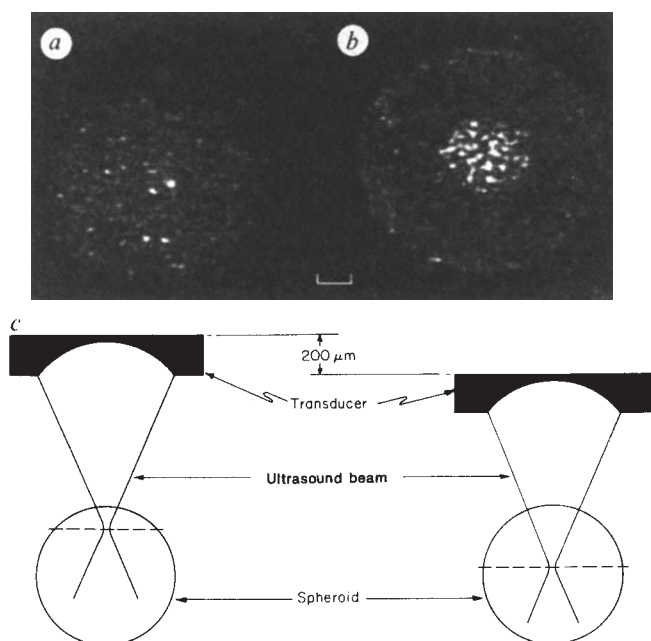


Fig. 2 Ultrasound microscope images of cross-sectional planes through *a* the cap and the center of an 850 µm diameter, MGH-U1 spheroid (*a* and *b* respectively). Scale bar, 100 µm. *c*, Scanning geometry employed for imaging.

Methods. Spheroids were grown according to the following regime. At day zero, 3×10^6 cells were placed in a 250 ml spinner flask (Johns Scientific, Toronto) containing 160 ml α -MEM (minimum essential medium) and 16 ml fetal calf serum with kanamycin (Sigma). Cells were maintained in a 37 °C water bath and spun at 110 r.p.m. Initially the cells formed aggregates (5–50 cells) which then grew by cell division to form spheroids. A spheroid, embedded in agar, was placed in a bath of phosphate buffered saline at room temperature and scanned with the ultrasound microscope.

which results in areas of hypoxia and eventually necrosis. The effects of oxygenation⁷, nutrition⁸, radiation⁹, and chemotherapeutic agents¹⁰ on the growth characteristics of cell population within spheroids have been studied to model the effect of therapy on tumours. There are a number of methods for assaying the growth characteristics of spheroid cells. Usually the spheroid is sectioned, fixed and stained for histological examination. But, the most useful assay is cell viability which involves disaggregating the spheroid and measuring the plating efficiencies of individual cells. The disadvantage of this technique is that cell viability may change before plating due to a loss of intercellular contact and changes in the biochemical environment¹¹. The use of a non-invasive technique like ultrasound backscatter microscopy may abrogate this problem.

Initially, we made images of an 850 µm diameter MGH-U1 spheroid derived from a human bladder carcinoma cell line (Fig. 2). The scanning geometry employed is shown in Fig. 2c. The grey scale represents the backscatter signal level, white corresponding to high backscatter and black to no backscatter. The speckled appearance of the images is characteristic of backscatter techniques and is due to a combination of structural information and coherent interference of ultrasound waves scattered from within a single resolution volume element. An image plane through the cap of the spheroid (Fig. 2a) shows a relatively homogeneous backscatter signal. However, the image of a plane closer to the centre of the spheroid (Fig. 2b), shows two distinct regions. The rim has a backscatter level similar to that of the cap, while the central region gives a surprising 20-fold (25 dB) increase in signal amplitude.

To understand the observed backscatter variations, ultrasound microscope images were made of cross-sectional planes at 300 µm depth intervals in a 1.5 mm diameter spheroid and compared to optical microscope images of histological sections

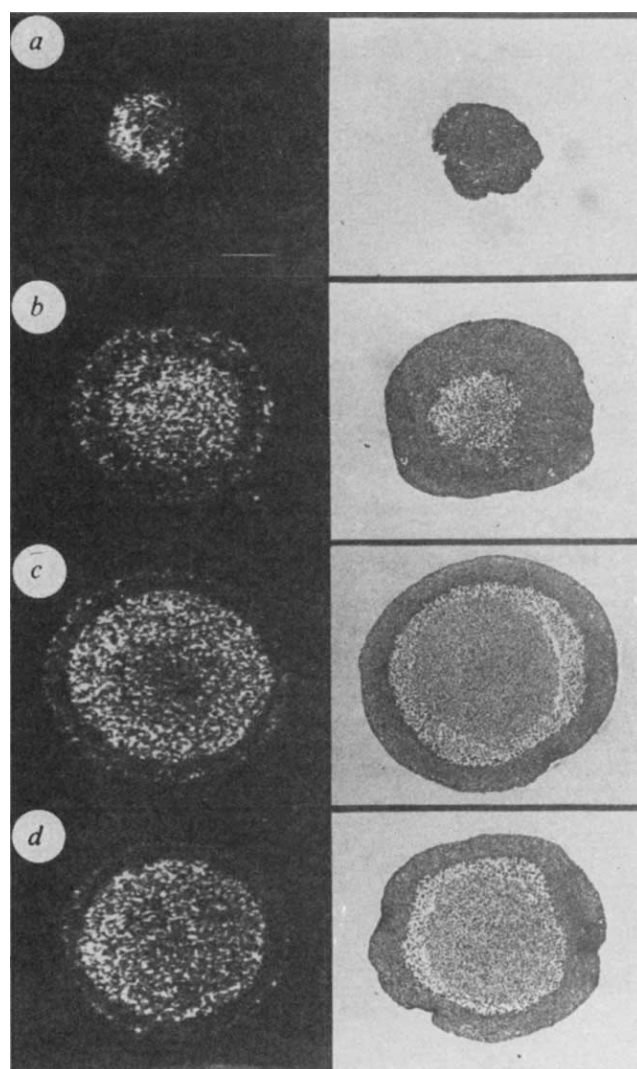


Fig. 3 Ultrasound microscope images of cross-sectional planes at 300 µm depth intervals in a 1.5 mm diameter, MGH-U1 spheroid (left) and optical microscope images of corresponding planes in the same spheroid (right). Scale bar, 300 µm.

Methods. Histological sections were prepared by first placing the small agar block containing the spheroid in Bouin's fixative for 24 h. The block was rinsed in double distilled water (DDW) and stored in 70% ethanol before being stained with 2% mercurchrome for 15 min. The block was again rinsed in DDW, set in 1.5% agar and stored in 10% formalin. The whole block was then dehydrated and set in paraffin for sectioning. Sections (10 µm thick) corresponding to the ultrasound image planes were cut and stained with haematoxylin and eosin and photographed under an optical microscope. The images were magnified to compensate for the 20% shrinkage that occurred during histological preparation.

of the same spheroid (Fig. 3). The ultrasound image through the cap of the spheroid (Fig. 3a) shows a signal level resulting from the homogenous cell mass observed in the optical microscope image. The large signal on the left of the ultrasound image is thought to result from specular reflection at the spheroid-agar interface. The ultrasound image of the central plane (Fig. 3c) clearly shows three distinct regions including a rim of width 150 µm which gives low backscatter, a 250 µm intermediate region of high backscatter and a central, 650 µm diameter region of reduced backscatter. The optical microscope image of this plane also shows three regions; a rim and intermediate region both of width 170 µm and a central region with a diameter of 800 µm.

The ultrasound backscatter levels from the three regions seen in Fig. 3c can be explained from observations of magnified

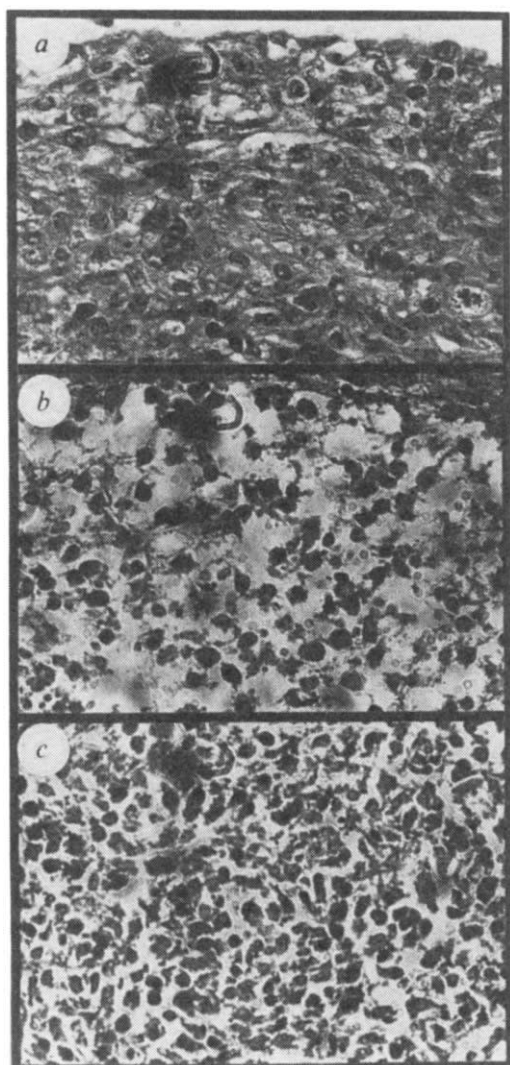


Fig. 4 Magnified optical microscope images of the rim (a), the intermediate (b) and central regions (c) of the spheroid observed in Fig. 3c (right). The backscatter variations in the ultrasound images can be explained by the variation in cellular composition through the spheroid.

optical images of these areas as shown in Fig. 4. The rim (Fig. 4a) consists of a viable cell mass and demonstrates tightly cohesive cells containing well defined cytoplasmic membranes and normal chromatin structure. A mitotic figure is evident in the lower right corner of this image. The low backscatter levels from this region are probably due to small acoustic impedance variations in the cell mass. The intermediate region (Fig. 4b) contains dead cells with disrupted cytoplasmic membranes. The nuclei are pyknotic and are surrounded by large spaces presumably containing the dispersed cytoplasmic fluid. Large acoustic impedance variations between the dense, collapsed nuclear chromatin and the surrounding liquid matrix would account for the high backscatter signals from this region. The central region (Fig. 4c) is similar to the intermediate region, except that less fluid is present. The distance between pyknotic nuclei is 2–3 μm . This is considerably less than an acoustic wavelength at 100 MHz (15 μm) so the central region appears more homogenous to the ultrasound wave than does the intermediate region. Consequently the central region gives a lower backscatter signal. The reduction in signal cannot be explained by attenuation effects because corresponding regions in Fig. 3b and d are not observed. Differences between the ultrasound and optical image measurements may be due either to the image planes not corre-

sponding exactly or to disproportionate shrinkage in the spheroid during processing for histological examination.

The spheroid system provided an appropriate starting point for the application of ultrasound backscatter microscopy because of its simple, well defined structure. Understanding the origins of the spheroid image features will aid the interpretation of more complicated images of intact tissue. In general, the contrast shown in these images indicates that this method of microscopy could be a useful and unique tool for subsurface microscopic imaging in living biological specimens. Further developments may allow imaging in real time (20 frames per second) and also improved resolution by the use of higher frequency transducers. Applications could then include the imaging of dynamic events at depth in living systems.

We thank Dr A. M. Rauth for suggesting the problem of spheroid imaging, D. Steele-Norwood for technical assistance, D. Booth for histological sectioning of the spheroids and Dr L. M. Sugar for analysing cellular composition in the sections. The work was supported by the National Cancer Institute of Canada and the Canadian Rhodes Scholars' Foundation.

Received 27 August; accepted 20 October 1987.

1. Lemons, R. A. & Quate, C. F. *Appl. Phys. Lett.* **24**, 163–165 (1974).
2. Lemons, R. A. & Quate, C. F. *Phys. Acoust.* **14**, 1–92 (Academic, New York, 1979).
3. Foster, J. S. & Rugar, D. *IEEE Trans. Ultrason. Acoust.* **32**, 139–151 (1985).
4. Forster, F. K., Olerud, J. E., Pomajevich, G. R., Holmes, A. W. & Sharar, S. R. *IEEE Ultrasonics Symp.* 957–962 (1986).
5. Sherar, M. D. & Foster, F. S. *Acoustical Imaging* vol. 16 (ed. L. W. Kessler) (Plenum, New York, in the press).
6. Hill, C. R. in *Physical Principles of Medical Ultrasonics* (ed. Hill, C. R.) 281–283 (Horwood, Chichester, 1986).
7. Mueller-Klieser, W., Freyer, J. P. & Sutherland, R. M. *Br. J. Cancer* **53**, 345–353 (1986).
8. Mueller-Klieser, W., Freyer, J. P. & Sutherland, R. M. *Adv. exp. Med. Biol.* **159**, 487–495 (1985).
9. Yuhas, J. M., Blake, S. & Weichselbaum, R. R. *Int. J. Radiat. Oncol. biol. Phys.* **10**, 2323–2327 (1984).
10. Sutherland, R. M., Eddy, H. A., Boreham, B., Reich, K. & Vanantwerp, D. *Cancer Clin. Trials* **3**, 78–83 (1980).
11. Sutherland, R. M. & Durand, R. E. *Curr. Topics Rad. Res.* **11**, 87–139 (1976).

DNA H form requires a homopurine–homopyrimidine mirror repeat

S. M. Mirkin*, V. I. Lyamichev*, K. N. Drushlyak*,
V. N. Dobrynin†, S. A. Filippov†
& M. D. Frank-Kamenetskii*

* Institute of Molecular Genetics, USSR Academy of Sciences, Moscow 123182, USSR

† M. M. Shemyakin Institute of Bioorganic Chemistry, USSR Academy of Sciences, Moscow, USSR

Regular homopurine-homopyrimidine tracts, $(dG-dA)_n$, $(dT-dC)_n$ and $(dG)_n$, $(dC)_n$, undergo a superhelix-induced, strongly pH-dependent, structural transition into a novel DNA conformation, the H form^{1–3}. We have suggested that the H form can arise in any homopurine-homopyrimidine mirror repeat (H palindrome)³. We have now tested this prediction using a tailored series of plasmids carrying the inserts AAGGGAGAAXGGGGTATAGGGGYAAGAGGGAA, where X and Y may be either A or G, and subject them to two-dimensional gel electrophoresis. In support of our hypothesis, the inserts exhibited facile transitions into the H form for X=Y=G, or X=Y=A, whereas the transition was much more difficult or impossible for the two non-palindromes (X=A, Y=G or X=G, Y=A). We present evidence that the H form is the structural basis for S1-nuclease hypersensitivity.

We have recently suggested that any mirror-repeated sequence of purines in one strand (H palindrome) would undergo a characteristic structural transition to the H form under superhelical stress³. An H palindrome may carry an arbitrary (not

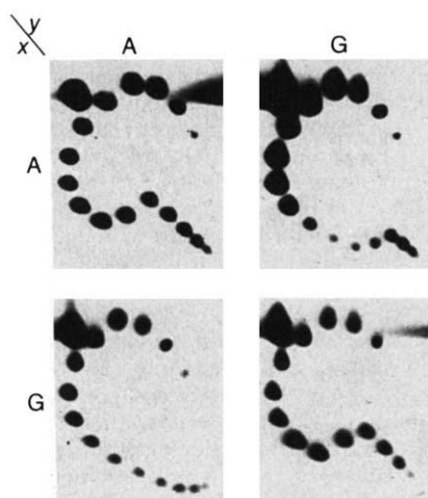
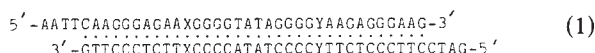


Fig. 1 2-D agarose gel electrophoresis patterns of four plasmids pXY32 at pH 4.3 under identical experimental conditions. Separation in the first direction (top to bottom) was in sodium citrate buffer (0.2 M in sodium) and chloroquine was added to relax superhelical stress in the second direction (left to right).

Methods. To tailor the plasmids we obtained four duplexes with cohesive ends (see text) using the phosphoramidite method and the automatic System 1 synthesizer (Beckman). Insert halves were synthesized and ligated in the desired combination with the vector DNA precut with *EcoRI*/*Bam*HI restriction endonucleases. These plasmids were used to transform *E. coli* JM83 cells. Sequencing the *EcoRI*/*Hind*III region of the pXY32 plasmids by the Maxam-Gilbert method confirmed that they carried the correct sequences.

necessarily homopurine) insert between the two halves of the repeat, though a region enriched with (dA·dT) pairs is preferred. To test this hypothesis we have synthesized a series of duplexes



where X, Y are purines in the top line or complementary pyrimidines in the bottom line; the flanking cohesive ends guaranteed incorporation between the *EcoRI* and *Bam*HI sites of the pUC19 polylinker. We term the resulting plasmids pXY32: the four variants being pAA32, pAG32, pGA32 and pGG32. As before^{1,3}, we use two-dimensional (2-D) gel electrophoresis to detect the transition to the H form.

If our concept is correct, we can expect supercoiling to readily induce a transition to the H form for the pAA32 and pGG32 plasmids, but the transition would be much more difficult or impossible for pAG32 and pGA32. Figures 1 and 2 show results to fulfil this prediction: the pAA32 and pGG32 plasmids carrying H palindromes undergo pH-dependent structural transitions like the regular H palindromes (dG-dA)_n·(dT-dC)_n and (dG)_n·(dC)_n studied previously^{1,3}, whereas the pAG32 and pGA32 plasmids transitions occur only at a much higher negative superhelicity. The pGA32 plasmid showed no transition even for the lowest pH values achieved. Note that for the cases having a detectable transition, the mobility drop corresponded to about three topoisomers, as expected. These data therefore provide evidence that the H palindrome is the sequence requirement for the H conformation of DNA to form.

The asymmetrical character of the 'matrix' in Fig. 1, namely different patterns for the pAG32 and pGA32 plasmids, is not unexpected. This is a consequence of the fact that mirror repeats,

like linguistic palindromes, are characterized by a pseudo-symmetry rather than by a true symmetry: they have the same sequence in both directions, one of which is correct (reading from left to right or 5' to 3' in DNA), whereas the other is 'incorrect'. Mirror repeats are thus fundamentally different from inverted repeats, which give rise to cruciforms and have a true dyad symmetry axis. In addition to the differential effect of the pseudo-symmetrical substitutions, the lack of true symmetry should result in different probabilities for the occurrence of the two 'isomers' of the H conformation, even for ideal H palindromes (see Fig. 3).

The data in Fig. 1 show that the two halves of the H palindrome interact with each other in the H form to give a spatial arrangement; this is inconsistent with an assumption that the acid form of homopurine-homopyrimidine regions is a modified duplex⁴. Our data also indicate that different 'non-canonical' (other than TAT and protonated CGC⁺) base triads have different effects on the H-form stability, which is not unexpected, considering the well known differential effect of base-pair mismatches on duplex stability.

The structure of homopurine-homopyrimidine tracts somehow renders them hypersensitive to S₁-nuclease¹⁻³. According to our model of the DNA H form^{2,3} (Fig. 3), this hypersensitivity arises from single-stranded regions inherent, albeit to different extents, in both homopurine and homopyrimidine strands. Moreover, the S₁ nuclease is an efficient probe for the H form because the pH optimum for this enzyme (pH 4.5) favours the H-form extrusion (see Fig. 2). Thus we suggest that the H form is the structural basis for hypersensitivity to S₁. Inspection of the localization of S₁-hypersensitive sites (SHS) showed that, although many SHS correspond to

Table 1 Irregular S₁-hypersensitive sites

Source	Position	Sequence	Ref.
chicken β-globin gene	-55	... GGGGAAGAGGAGGGG ...	5, 6
rat preproinsulin II gene	-50	... GGGGTCAGGGG ...	6
SV40 enhancer (72 bp repeat)	—	... GGGGAGCCTGGGG ...	6
<i>Drosophila</i> heat shock genes:			
hsp 26	-100	... AGAGAGAGAAGAGAAGAGAGAGA ...	7
hsp 22	-340	... TCTCTCCCACTCTCT ...	8
hsp 70, pHΔ5'	-130	... CTCTCTGTACTATTGCTCTCTC ...	8
hsp 70, pH1	-150	... CTCTCTTTTCTTTTGGGTCTCTC ...	8
adenovirus major late promoter for different serotypes:			
Ad2	-30	... GGGGGCTATAAAAGGGGG ...	8-11
Ad7	-30	... GGGGGTATAAAAGGGGG ...	11
Ad12	-30	... GGGCTATAAAAGGG ...	11
human c-myc gene	-300	... CCCTCCCCATAAGCGCCCCCTCCC ...	12
human α2-α1 globin intergenic region	-1090	... CCTCCTCCACCTCCTCC ...	13
chicken α2(I) collagen promoter	-200	... TCCCTCCCCCTTCCTCCCTCCCT ...	14, 15

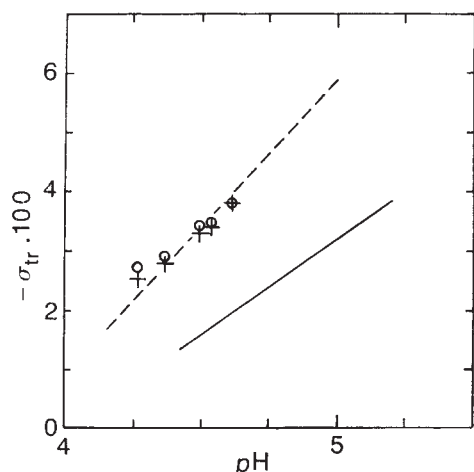


Fig. 2 Superhelix density of transition (σ_{tr}) as a function of pH. (—) pEJ4 plasmid carrying the (GA)₁₆ region¹; (---) pG31 plasmid carrying the (G)₃₁ region³; (○) pAA32 plasmid; (+) pGG32 plasmid.

long regular homopurine-homopyrimidine regions, a large number of them do not. Virtually all of them, however, proved to be H palindromes (see Table 1), so it is tempting to interpret the SHS presented in Table 1 in terms of the H form. It could be argued that in some cases the palindromic regions in Table 1 are too short to form stable triplexes. Very short inverted repeats cannot form cruciforms^{16,17}, but it should be emphasized that the situation with the H form is quite different from the case of the cruciform in that the CGC⁺ base-triads become extremely stable as the pH decreases. As this triad is stabilized by the proton, its stability constant is^{1,18}:

$$s = s_0(1 + 10^{pK-pH})$$

where s_0 is the stability constant of the non-protonated CGC base-triad (when $pH > pK$). Values for s_0 and pK are still not accurately known but available data^{19,20} indicate that $pK \sim 8$ and $s_0 \sim 10^{-2}$. Thus at pH near 4, the s value is roughly estimated as 100. For Watson-Crick base pairs, s cannot be made more than about 10 by changes in ambient conditions. We therefore believe that the unusual stability of the triplex is attributable to the formation of the H form at low pH and at high negative superhelicity, even for very short H palindromes. Probably the H-form extrusion could not have been observed by 2-D gel electrophoresis for the H palindromes listed in Table 1, though a structural transition is clearly detectable by enzymic probing. We explain this apparent contradiction in the following way. First, an H palindrome must be long enough to observe a clear-cut mobility drop. Second, enzymatic or chemical probing give a positive result even if only a tiny proportion of DNA molecules in the sample have transiently adopted the H form, because only cut molecules are detected when radiolabel is used.

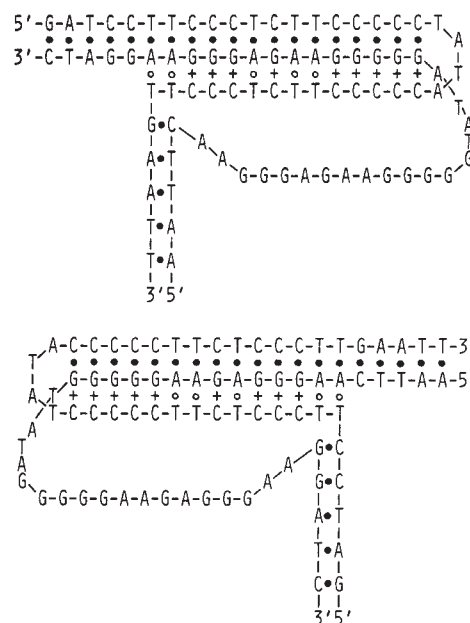


Fig. 3 The H-form structure of the pGG32 plasmid. The major element of the structure is a triplex which includes the Watson-Crick duplex (●) associated with the homopyrimidine loop through Hoogsteen base pairing (○, +). The two possible 'isomeric' forms are shown. Although these two forms are energetically different because of symmetry considerations (see text), the available data cannot discriminate between them.

Elsewhere we present direct experimental evidence to the effect that chemical and enzymic methods are much more sensitive probes of the H form than 2-D gel electrophoresis (Voloshin *et al.*, in preparation).

Palindromes are found at the replication origin of small DNA tumour viruses such as polyoma and SV40 (refs 21, 22) which are mirror repeat sequences having a common motif, CCTC(N)_nCTCC, and an AT-enriched central region and are therefore typical of an H-palindrome. In SV40 this sequence serves as a binding site of the large T antigen. The binding of the T antigen at this site involves a structural transition in DNA^{23,24}, which has been interpreted in terms of DNA bending. Our data suggest the alternative interpretation that the H palindromes at the replication origin of small tumour viruses form the H form. The T antigen could be an H-binding protein which converts the H-palindrome into the H form on binding. Such an H-binding protein could induce short H palindromes to undergo transition to the H form under physiological conditions of neutral pH.

We thank G. Malenkov and S. Nedospasov for advice, and Y. Lazurkin and A. Vologodskii for discussion.

Received 26 May; accepted 1 October 1987.

- Lyamichev, V. I., Mirkin, S. M. & Frank-Kamenetskii, M. D. *J. biomolec. Struct. Dyn.* **3**, 327-338 (1985).
- Lyamichev, V. I., Mirkin, S. M. & Frank-Kamenetskii, M. D. *J. biomolec. Struct. Dyn.* **3**, 667-669 (1986).
- Lyamichev, V. I., Mirkin, S. M. & Frank-Kamenetskii, M. D. *J. biomolec. Struct. Dyn.* **5**, 275-282 (1987).
- Pulleyblank, D. E., Hanford, D. B. & Morgan, A. R. *Cell* **42**, 271-280 (1985).
- Schon, E., Evans, T., Welsh, J. & Efstratiadis, A. *Cell* **35**, 837-848 (1983).
- Evans, T., Schon, E., Gora-Maslak, G., Patterson, J. & Efstratiadis, A. *Nucleic Acids Res.* **12**, 8043-8058 (1984).
- Siegfried, E., Thomas, G. H., Bond, U. M. & Elgin, S. C. R. *Nucleic Acids Res.* **14**, 9425-9444 (1986).
- Mace, H. A. F., Pelham, H. R. B. & Travers, A. A. *Nature* **304**, 555-557 (1983).
- Goding, C. R. & Russell, W. C. *Nucleic Acids Res.* **11**, 21-36 (1983).
- Yu, Y.-T. & Manley, J. L. *Cell* **45**, 743-751 (1986).
- Kilpatrick, M. W., Torri, A., Kang, D. S., Engler, J. A. & Wells, R. D. *J. biol. Chem.* **261**, 11350-11354 (1986).

- Boles, T. C. & Hogan, M. E. *Biochemistry* **26**, 367-376 (1987).
- Shen, C.-K. *J. Nucleic Acids Res.* **11**, 7899-7910 (1983).
- McKoen, C., Schmidt, A. & de Crombrughe, B. *J. biol. Chem.* **259**, 6636-6640 (1984).
- Finer, M. M., Fodor, E. J. B., Boedtker, H. & Doty, P. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1659-1663 (1984).
- Vologodskii, A. V., Lukashin, A. V., Anshelevich, V. V. & Frank-Kamenetskii, M. D. *Nucleic Acids Res.* **6**, 967-982 (1979).
- Vologodskii, A. V. & Frank-Kamenetskii, M. D. *FEBS Lett.* **143**, 257-260 (1982).
- Lyamichev, V. I., Mirkin, S. M., Frank-Kamenetskii, M. D. & Cantor, C. R. *Nucleic Acids Res.* (in the press).
- Howard, F. B., Frazer, J., Lippset, M. N. & Miles, H. T. *Biochem. biophys. Res. Commun.* **17**, 93-102 (1964).
- Lee, J. S., Johnson, D. A. & Morgan, A. R. *Nucleic Acids Res.* **6**, 3073-3091 (1979).
- Soeda, E., Arrand, J. R., Smolar, N. & Griffin, B. E. *Cell* **17**, 357-370 (1979).
- Riley, D. E., Reeves, R. & Gartler, S. M. *Nucleic Acids Res.* **14**, 9407-9423 (1986).
- Ryder, K. *et al. Cell* **42**, 539-548 (1985).
- Ryder, K., Silver, S., DeLucia, A. L., Fanning, E. & Tegtmeyer, P. *Cell* **44**, 719-725 (1986).

Patch-clamping of the inner mitochondrial membrane reveals a voltage-dependent ion channel

M. Catia Sorgato, Bernhard U. Keller* & Walter Stühmer*

Istituto di Chimica Biologica e Centro CNR per lo Studio della Fisiologia Mitochondriale, I-35131 Padova, Italy

* Max-Planck-Institut für biophysikalische Chemie, Abt. Membranbiophysik, D-3400 Göttingen, FRG

The prime function of mitochondria is to provide the cell with adenosine triphosphate (ATP). ATP synthesis is driven by the protonmotive force (Δp), which is generated and maintained across the inner mitochondrial membrane (IMM) by the activity of the respiratory chain¹. It is widely believed that the IMM is unlikely to contain ion channels like those present in the plasma membrane, because the high rates of ion transport characteristic of open channels would be expected to dissipate the Δp . Although the small size of the organelle has prevented the use of classical electrophysiological methods, the recent introduction of the patch-clamp technique, which allows currents to be recorded from very small cells², has enabled us to test this hypothesis. By patch-clamping the IMM, we have identified a slightly anion-selective channel, which is voltage-dependent and has a mean conductance of 107 pS in the presence of symmetrical 150 mM KCl.

In order to patch-clamp the IMM, organelles had to be of suitable size and it was necessary to remove the outer mitochondrial membrane (OMM). Therefore, we isolated giant mitochondria from cuprizone-fed mice³ and removed the OMM using a swelling-shrinking-sonication procedure⁴. This yielded vesicles (mitoplasts) with diameters of 3–6 μm . Electron micrographs of these mitoplasts (Fig. 1a) show that the OMM is not always completely removed but may remain attached to the IMM in discrete regions, the mitoplast cap (compare with the whole OMM of parent mitochondria, Fig. 1b). When observed under

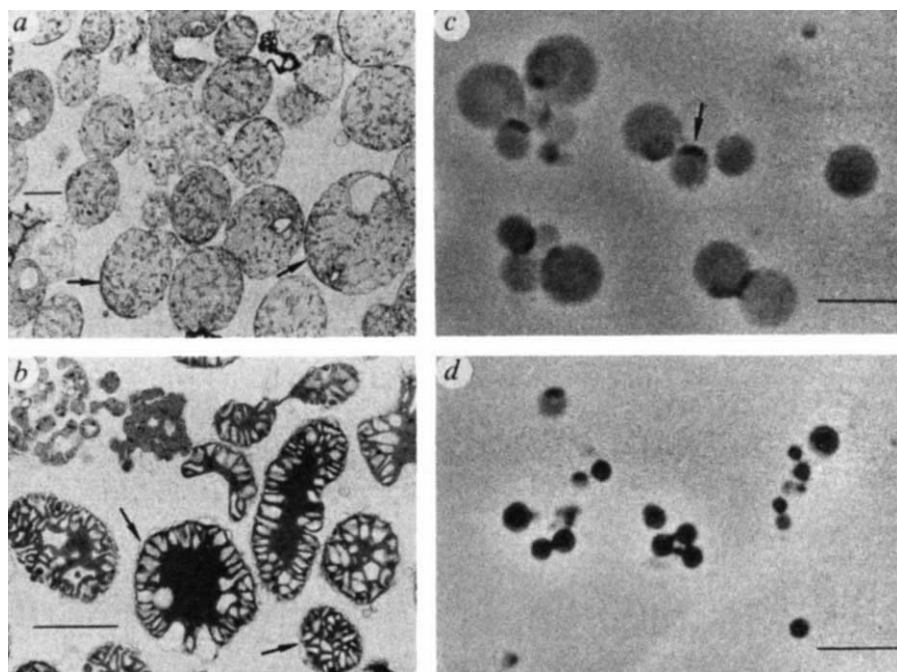
phase contrast, the cap region is clearly distinguishable (Fig. 1c). Care was taken not to place the patch electrode on the cap. It was also easy to avoid contaminating parent mitochondria, which appear smaller and denser under phase contrast (Fig. 1d). The functional integrity of the whole mitochondria was ascertained with a tetraphenylphosphonium-selective electrode⁵. In such measurements, giant mitochondria were able to maintain a membrane potential of about 150 mV.

To investigate the ionic currents through the IMM, mitoplasts were patch-clamped². Figure 2a shows the steady-state current-voltage relationship ($I-V$) recorded from a whole mitoplast in a patch-clamp configuration analogous to the whole-cell configuration². For negative membrane potentials, the IMM displays a high resistance and the membrane current is small. For positive voltages the curved $I-V$ indicates a high ionic permeability. In subsequent experiments we could identify this permeability at the single channel level. In mitoplast-attached patches, we observed current from single channels which tend to close at negative potentials and open at positive potentials (Figs 2b, 3). Therefore, we propose that the large whole-mitoplast current in Fig. 2a corresponds to the macroscopic current flowing through single channels (Fig. 2b). This is further supported by the pulse-protocol displayed in Fig. 3a. Here, at least four IMM channels in a mitoplast-attached patch open for membrane depolarizations to +60 mV. The opening and closing of the IMM channels at that voltage is shown in Fig. 3b on an expanded timescale. Fast events in channel gating are best seen in Fig. 3c, where a single channel record is shown at different voltages under steady-state conditions. Figure 3d presents the IMM channel amplitude distribution, at +60 mV, fitted by a gaussian. We find a conductance of 107 ± 8 pS ($n=11$) over a range of -30 mV to +80 mV.

To determine the ionic selectivity of the channel, the reversal potential of the single channel current was measured using either mitoplast-attached or excised configurations. In a bath solution of 150 mM KCl and a pipette solution of 30 mM KCl, the reversal potential was -22 mV (pipette potential), giving a permeability ratio of 4.5 for chloride to potassium. The effect of protons was assessed by using symmetrical 150 mM KCl solution

Fig. 1 Giant mitoplasts (a, c) and mitochondria (b, d) visualized with electron (a, b) and phase contrast (c, d) microscopy. The arrows in a, b and c point to the residual outer membrane (cap) in mitoplasts and to the integral outer membrane in mitochondria. Scale bars: a and b, 1 μm ; c and d, 10 μm .

Methods. For electron microscopy, an aliquot of isolated mitoplasts or mitochondria, suspended in 250 mM mannitol, 50 mM HEPES-KOH (pH 7.2), was fixed with four aliquots of 1.25% glutaraldehyde (in the suspending medium) and postfixed (2 h at 4 °C) with 1% osmium tetroxide (in 0.12 M sodium cacodylate, pH 7.2). The sample was then *en bloc* stained (4 h at 4 °C) with 1% uranylacetate (in 0.12 M sodium acetate, pH 5.5), dehydrated in ethanol and embedded in EPON 812. Thin sections were cut, examined and photographed under the electron microscope. For phase contrast microscopy, an aliquot of mitoplasts or mitochondria was diluted with 150 mM KCl, 20 mM HEPES-KOH (pH 7.2), 0.1 mM CaCl_2 and photographed with a Zeiss photo microscope. Giant mitochondria and mitoplasts were obtained as described previously³, except that, after the second centrifugation, the interphase between the two layers (and not the pellet) was resuspended in the 0.5 M sucrose (lower) layer. Although sonication of the vesicles following the swelling-shrinking step was occasionally performed, this was usually omitted as electron micrographs of the sonicated vesicles showed that also in this population fragments of the outer membrane remained attached to the IMM. The mitoplast pellet was always washed a second time with 250 mM mannitol, 50 mM HEPES-KOH (pH 7.2).



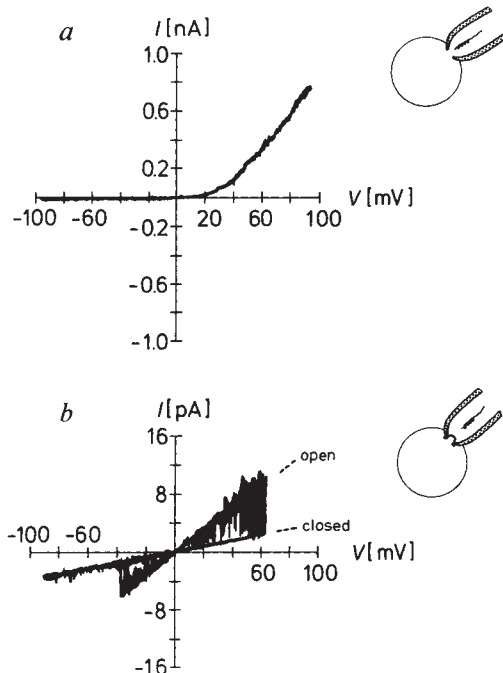


Fig. 2 Current-voltage relation of mitoplasts under different recording conditions. *a*, Steady state $I-V$ in whole mitoplast configuration. *b*, $I-V$ of a mitoplast-attached patch, showing single channel gating. The two diagonals (*b*) correspond to the closed (flat diagonal) and open state (steep diagonal) of the channel. Both records were obtained by repeatedly manually sweeping the potential in the pipette. In *a*, the rate of potential change was slow enough as to achieve steady state conditions. Symmetrical solutions of 150 mM KCl were used for either record. Note that all potentials are given according to the standard electrophysiological convention.

Methods. Aliquots (5–10 μ l) of mitochondria or mitoplasts, obtained as described in Fig. 1, were layered on the bottom of a chamber and then diluted with 200–300 μ l of bath solution: usually 150 mM KCl, 0.1 mM CaCl_2 , 20 mM HEPES-KOH (pH 7.2). After allowing the vesicles to adhere to the bottom of the chamber (minimum 15 min), the chamber was mounted on the patch-clamp setup² and extensively perfused with the bath solution. The pipette was usually filled with the same solution as that in the bath. Occasionally, CaCl_2 was omitted and this did not alter the single channel activity. In some experiments, the KCl concentration of either the bath or the pipette solution was varied. If they differed in ionic content, the osmolarity of the lower ionic strength solution was adjusted with sucrose to that of the higher ionic strength solution (305 mosM). Presence of 200 μ M quinine¹², either in the bath or in the pipette solution did not block the activity of the channel. We found by reversal potential measurements that the inner solution of the mitoplasts equilibrated with the bath solution in less than 30 min. Decreasing the osmolarity of both pipette and bath solutions (50 mM KCl, 0.1 mM CaCl_2 , 1 mM HEPES-KOH (pH 7.2)), reduced the amplitude of the single channel currents but otherwise had no additional effect. Pipettes of 5–30 M Ω resistance were used. Seals with resistances greater than 10 G Ω formed relatively easily, providing the chamber had been extensively perfused. In some experiments, whole-mitoplast patches occurred spontaneously, as indicated by an increase in membrane capacitance. Experiments were carried out at room temperature.

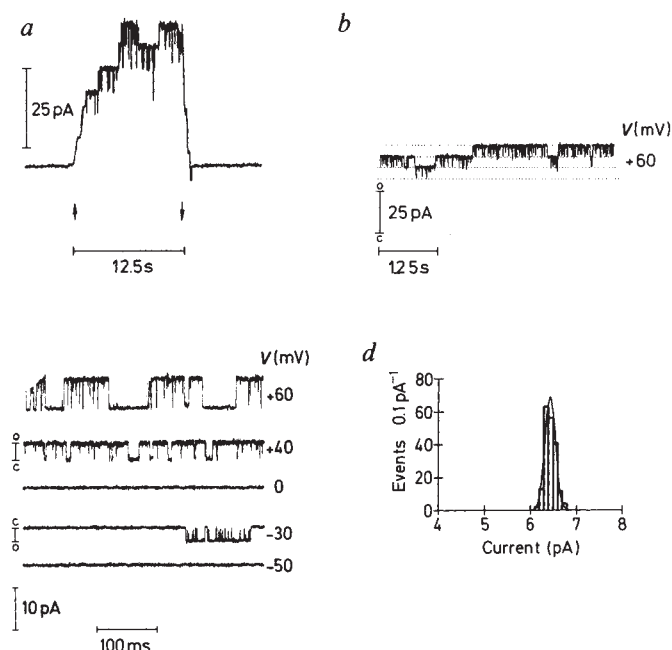


Fig. 3 Single channel records of the IMM channel. *a*, Single channel currents in response to a depolarizing membrane voltage pulse. The trace shows ionic currents through at least four IMM channels in a mitoplast-attached patch, in symmetrical 150 mM KCl. A depolarization from -30 to $+60$ mV was applied during the time elapsed between arrows. Note the short opening of a membrane channel after membrane repolarization to -30 mV. Current records were stored on a RACAL tape recorder, filtered at 1 kHz (8-Pole Bessel filter) and stored on a computer for further analysis and graphical display. Current fluctuations at positive voltages correspond to rapid openings and closings of the membrane channels. *b*, *c*, Records equivalent to Fig. 3*a* on a 10-fold (*b*) and 125-fold extended timescale (*c*) reveal fast and slow components in the gating kinetics of the channel. At $+60$ mV (upper record) the distribution of open times is well approximated by a single exponential with a time constant $t = 6.9$ ms. The closed time distribution is characterized by a fast process ($t = 1.9$ ms), with an additional slow component ($t > 100$ ms). At negative membrane potentials channel closing dominates the gating process under steady-state conditions (-30 mV and -50 mV in *c*). *d*, Amplitude histogram of the IMM channel. Records like those displayed in Fig. 3*c* were obtained at $V = +60$ mV and analysed with the computer program TAC¹⁸. The measured amplitude distribution is well approximated by a gaussian with mean (a) = 6.4 pA and s.e.m. = 0.1 pA (solid line), giving a conductance value of 106.6 ± 8 pS (reversal potential ± 2 mV).

and varying the pH between 6.2 and 9.0, in either the bath or the pipette solution, while keeping the pH of the other solution fixed at 7.2. Under these conditions the reversal potential was consistently found to be close to 0 mV and no effect on the gating kinetics was observed. Thus, over this pH range, protons have little effect on either the amplitude or gating of the single channel currents, ruling out the possibility that either the internal or external pH of the mitoplasts functions as a physiological regulator of the channel.

Interestingly, the properties of the IMM channel described here are very different from those of the voltage-dependent anion channel (VDAC) of the OMM, which has been characterized after reconstitution in planar lipid bilayers⁶. That the IMM channel is a separate channel type was further substantiated by patch-clamping whole mitochondria. In these studies we were able to record currents through single channels with a conductance of approximately 350 pS in symmetrical 150 mM KCl (Fig. 4). Taking into account the difference in membrane environment,

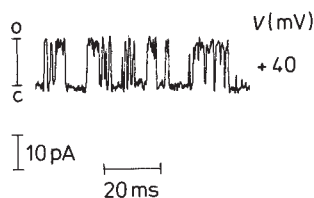


Fig. 4 Channel from OMM. Single channel currents were measured in symmetrical 150 mM KCl, from a mitochondria-attached patch, at +40 mV. These channels display a mean conductance of 347 ± 26 pS.

this value might correspond to the single channel conductance of 480 pS (100 mM KCl), found for the VDAC after reconstitution in planar lipid bilayers⁷.

To our knowledge, this is the first report of patch-clamp studies of the IMM (the patch-clamping of the OMM was recently reported⁸), and of the presence of a high conductance ion channel in it. The possible existence of an inner membrane protein, able to translocate anions electrophoretically, has already been suggested⁹⁻¹². Unfortunately the turnover of this anion channel is not yet known, though it is known to be maximally activated at alkaline pH and to be regulated also by magnesium and calcium ions. These characteristics and the lack of effect of quinine¹² on the channel activity described here render it unlikely that they are the same channel. But until comparable data are available, and because of the different incubation and experimental conditions used by us, we cannot definitively rule out this possibility.

It is important to understand the possible physiological function and regulation of the IMM channel. The presence of the channel in most membrane patches, together with the high conductivity displayed, suggests a non-trivial role in the regulatory processes of the IMM and of the mitochondrion in general, many of which are not yet fully appreciated or understood^{12,13}. Indeed, with the particular properties of the IMM component described here, several functions could be feasible. This channel could be regarded as an intrinsic uncoupling protein analogous to, and with similar function to, that of brown fat mitochondria¹⁴. Thus, the IMM channel could be of key importance in providing heat in tissues other than brown fat.

Mitochondria are known to undergo finely tuned volume changes¹⁵. In line with that envisaged for the anion channel¹², the IMM channel could represent a safeguard mechanism for the maintenance of volume homeostasis of mitochondria. Such control is extremely important under normal but especially under pathological conditions, such as ischemia¹², where the excessive swelling of the matrix impairs those mitochondrial functions essential for cell survival. Undoubtedly the high ion conductance of the IMM channel renders it highly suitable for such a role.

Finally, another tantalizing possibility, which involves mitochondrial biogenesis, must be taken into consideration. The majority of mitochondrial proteins are synthesized in the cytoplasm and hence have to be imported into the organelle via a mechanism which is not known in detail¹³. Hurt and van Loon¹⁶ have posed the question of whether channels can mediate such transport. As a proteinaceous component of the OMM has recently been found to be part of the import apparatus¹⁷, it will be of interest to ascertain whether the IMM has a similar role and if it is involved in the continuous traffic of macromolecules occurring between the organelle and the cytoplasm¹³.

M.C.S. thanks EMBO for the award of a fellowship. We thank Dr Geoffrey Fox for the electron micrographs, Dr Michel Robert-Nicaud for the phase contrast pictures and Drs Erwin Neher, Frances Ashcroft and Meyer B. Jackson for comments on the manuscript.

Received 13 August; accepted 13 October 1987.

1. Mitchell, P. *Chemiosmotic Coupling in Oxidative and Photosynthetic Phosphorylation* (Glynn, Bodmin, Cornwall, 1966).
2. Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. *Pflügers Arch. ges. Physiol.* **391**, 85-100 (1985).
3. Hochman, J., Ferguson-Miller, S. & Schindler, M. *Biochemistry* **24**, 2509-2516 (1985).
4. Sottocasa, G. L., Kuylensstierna, B., Ernster, L. & Bergstrand, A. *Meth. Enzym.* **10**, 448-463 (1967).
5. Ferguson, S. J. & Sorgato, M. C. A. *Rev. Biochem.* **51**, 185-217 (1982).
6. Colombini, M. *Nature* **279**, 643-645 (1979).
7. Roos, N., Benz, R. & Brdiczka, D. *Biochim. biophys. Acta* **686**, 204-214 (1982).
8. Tedeschi, H., Mannella, C. A. & Bowman, C. L. *J. membrane Biol.* **97**, 21-29 (1987).
9. Azzi, A. & Azzone, G. F. *Biochim. biophys. Acta* **120**, 466-468 (1966).
10. Brierley, G. P. *Biochem. biophys. Res. Commun.* **35**, 396-402 (1969).
11. Selwyn, M. J., Dawson, A. P. & Fulton, D. V. *Biochem. Soc. Trans.* **7**, 216-219 (1979).
12. Garlid, K. D. & Beavis, A. D. *Biochim. biophys. Acta* **853**, 187-204 (1986).
13. Yaffe, M. & Schatz, G. *Trends biochem. Sci.* **9**, 179-181 (1984).
14. Nicholls, D. G. *Biochim. biophys. Acta* **549**, 1-29 (1979).
15. Tedeschi, H. & Harris, D. L. *Archs biochem. Biophys.* **58**, 52-67 (1955).
16. Hurt, E. C. & van Loon, A. P. G. M. *Trends biochem. Sci.* **11**, 204-207 (1986).
17. Pfanner, N. & Neupert, W. *J. biol. Chem.* **262**, 7528-7536 (1987).
18. Sigworth, F. in *Single Channel Recording* (eds Sakmann, B. & Neher, E.) 91-105 (Plenum, New York, 1983).

ERRATUM

A giant intergalactic H I bubble near Arp143

P. N. Appleton, F. D. Ghigo, J. H. van Gorkom, J. M. Schombert & Curtis Struck-Marcell

Nature **330**, 140-142 (1987).

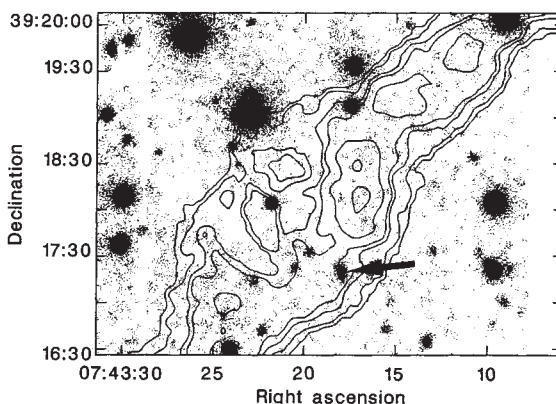
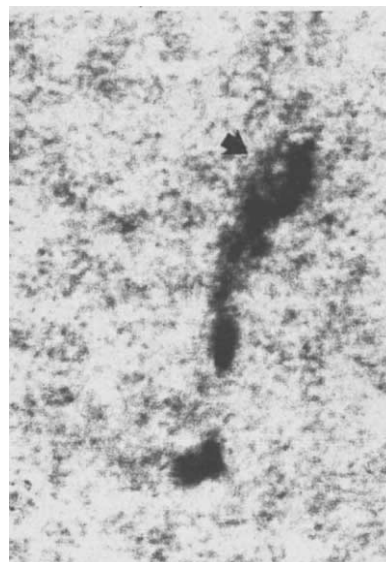


Fig. 1b

Fig. 2



In this paper the scale on Fig. 1b was printed incorrectly and Fig. 2 was shown upside down through no fault of the authors. The two figures are correct as printed here.

T cell antigen receptor antibodies

P. Kung

Immunology has a long tradition of using the immune system to study the immune system. Monoclonal antibodies are now being focused upon the antigen receptor of the T lymphocyte.

T LYMPHOCYTES have antigen receptors that are structurally similar to the well-characterized antigen receptors on B lymphocytes. The T cell receptor is a heterodimer having variable (V) regions and constant (C) regions, each of which is encoded in gene segments. The gene segments of the V regions rearrange to form the mature receptor chains during ontogeny. The great diversity of antigen specificities arises from the variety of choices for the V section segments, as well as from the highly variable joints between the segments.

Beyond this, the similarities between B and T lymphocyte antigen receptors fade rapidly. The T cell antigen receptor exists as a complex with at least three other proteins known as the CD3 (T3) complex (Fig. 1). The name for this complex is derived from the monoclonal antibody OKT3 developed by Ortho Pharmaceuticals (Raritan, New Jersey), that was shown through flow cytometry to mark T cells selectively. The functional significance of the CD3 complex became apparent when it was found that OKT3 is a strong mitogen when added to cultures of peripheral blood mononuclear cells¹. The suspicion that the CD3 complex has something to do with the antigen receptor was verified by biochemical characterization of the complex. Immunoprecipitation with OKT3 using non-ionic detergents revealed three proteins (γ , δ , ϵ) of the CD3 complex as well as a clonotypic heterodimer that has been shown to be the α and β chains of the T cell receptor. More recently, anti-clonotypic antibodies were raised against the α and β chains, some of which also exhibit mitogenic properties in lymphocyte cultures.

The antibody quest

The desire to study and manipulate the mechanism of T cell activation has led to attempts to generate a monoclonal antibody that interacts with the cell surface $\alpha\beta$ heterodimer of all T cells. To date, attempts to generate a monoclonal antibody that binds unequivocally to an invariant defined region of the cell surface $\alpha\beta$ heterodimer (as opposed to a closely associated protein of the CD3 complex) have failed. These attempts have resulted in monoclonal antibodies that detect clonotypic or variable region subset-specific determinants. In the murine system, a monoclonal antibody, KJ16-133, has been generated to an interstrain variation in the T cell β -variable region reper-

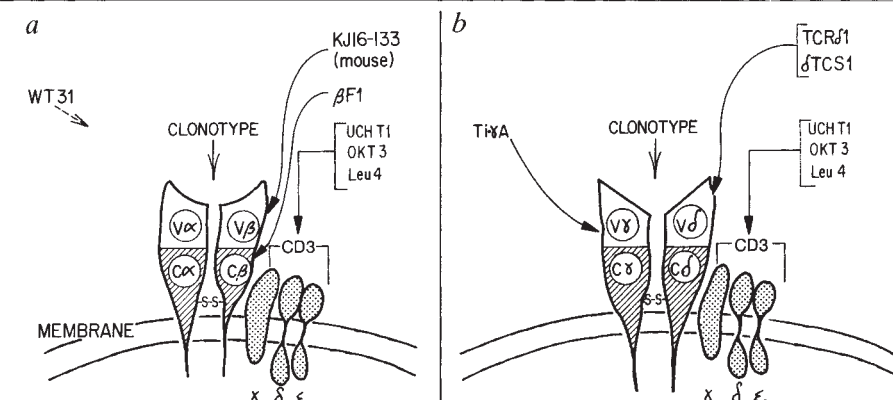


Fig. 1 Monoclonal antibodies defining the T cell receptor. *a*, β F1 and WT31 identify $\alpha\beta$ cells, and β F1 detects all β -chains. *b*, δ TCS1 and anti-TCR δ 1 identify $\gamma\delta$ cells, and Ti- γ A identifies a large subset of the $\gamma\delta$ cells. The CD3 complex is common to both receptors. The newly described ξ and p21 components¹² have been omitted for clarity. Clonotypic determinants are limited to a single cell clone. Courtesy C. Maplethorpe.

toire². This monoclonal antibody detects one family of variable region genes known as $V_{\beta}8$, similar to the classical anti- V_H allotype reagents that have been used to study rabbit immunoglobulin variable region genes. Some of the ostensibly "anti-clonotypic" monoclonal antibodies being generated against human T cells will undoubtedly turn out to detect variable region families, and permit the manipulation of small populations of T cell repertoire subsets. The application of these monoclonals to diseases with T cell involvement is readily apparent. For example, an experimental murine T cell tumour is growth-arrested by an anti-clonotypic monoclonal antibody directed against its receptor³.

The great difficulty in generating a reagent for the detection of the constant region of the $\alpha\beta$ T cell receptor led to the notion that these determinants are physically masked by the proteins of the closely associated CD3 complex. Such determinants are exposed by the denaturing conditions that are used in techniques such as immunoprecipitation, immunoblotting, or fixed-tissue staining, and screening procedures based upon these techniques resulted in the discovery of the first monoclonal antibody directed against the human T cell receptor β -chain constant region, β F1 (ref. 4; the antibody is available from T Cell Sciences). This reagent detects non-CD3-associated β -chain in the cytoplasm, as well as β -chain on the surface of fixed cells. The reactivity of this reagent allows the study of β -chain expression in normal lymphocyte development, and permits the classi-

fication of leukaemias and lymphocytes according to these stages⁵.

The suspected existence of a small subset of T cells that do not express $\alpha\beta$ T cell receptors, so-called $\gamma\delta$ cells, is given strong support by the demonstration that a small subset of CD3-positive cells do not express the β -chain protein. This finding supports the observation that another $\alpha\beta$ cell-specific monoclonal antibody, WT31⁶, detects fewer T cells than does OKT3; however, WT31 does not readily immunoprecipitate nor does it react in a Western blot, leaving the epitope recognized by WT31 unknown. Reports that the proteins of the CD3 complex from $\alpha\beta$ cells migrate differently on gels than do those from $\gamma\delta$ T cells⁷ do not lessen the mystery of the WT31 determinant, although they support its use as a flow cytometry reagent helpful in differentiating the two T cell types.

Two monoclonal antibodies, anti-TCR δ 1⁸ and δ TCS1⁹, identified by our company, react with the δ -chain of the human $\gamma\delta$ T cell receptor. Both anti-TCR δ 1 and δ TCS1 recognize all $\gamma\delta$ cell lines that have been tested, a feat not yet duplicated by monoclonals directed against $\alpha\beta$ cell lines. δ TCS1 was identified by screening for down-modulation of the CD3 complex on $\gamma\delta$ cell lines (while not affecting $\alpha\beta$ cell lines), followed by immunochemical characterization of the antigen. This procedure ensures that δ TCS1 not only detects the intact $\gamma\delta$ receptor, but also has mitogenic activity. δ TCS1 provides a strong specific stimulus in cultures of peripheral blood mononuclear cells, causing the rapid outgrowth

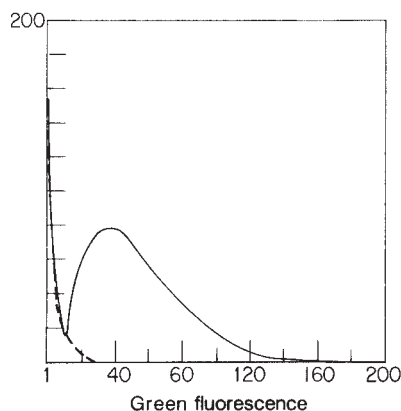


Fig. 2. δ TCS1 stimulates $\gamma\delta$ cells to proliferate and grow. Dotted line: flow cytometric profile of peripheral blood mononuclear cells (PBM) using FITC- δ TCS1 before starting cell culture. Solid line: flow cytometric profile of the same PBMs using FITC- δ TCS1 after four weeks' stimulation with δ TCS1. Mean doubling time was 36 hours. Courtesy C. Rittershaus.

of $\gamma\delta$ cells (Fig. 2). A third monoclonal antibody, Ti- γ A¹⁰ is specific for the human γ -chain. Ti- γ A detects a large subset of the $\gamma\delta$ cells in peripheral blood.

The apparent ease with which monoclonals with "anti-constant region" activity in the $\gamma\delta$ T cell subset have been generated has been explained by the molecular characterization of the $\gamma\delta$ T cell repertoire. The human γ variable region locus has only eight potentially-expressed genes detected to date¹¹, and the human δ variable region may have only one⁸. Since both anti-TCR δ 1 and δ TCS1 recognize the δ -chain, they behave functionally as anti-constant region reagents. The development of a panel of monoclonal antibodies that will completely define the $\gamma\delta$ T cell receptor repertoire should be an attainable goal.

Monoclonal antibodies are opening the door on the new frontier in immunology. No one knows the nature of the ligand for the $\gamma\delta$ T cell receptor, or what these cells do. But with new reagents, these questions may soon be answered. □

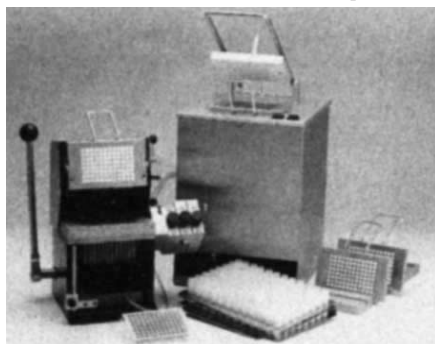
Patrick Kung is at T Cell Sciences, Inc., 840 Memorial Drive, Cambridge, Massachusetts 02139, USA. For more information, fill in reader service number 100.

1. Van Wauwe, J.P., de Mey, J.R. & Goossens, J.G. *J. Immunol.* **124**, 2708-2713 (1980).
2. Haskins, K. *et al. J. exp. Med.* **160**, 452-471 (1984).
3. Ashwell, J.D., Longo, D.L. & Bridges, S.H. *Science* **237**, 61-64 (1987).
4. Brenner, M.B. *et al. J. Immunol.* **138**, 1502-1509 (1987).
5. Picker, L.J., Brenner, M.B., Weiss, L.M., Smith, S.D. & Warnke, R.A. *Am. J. Path.*, in the press.
6. Tax, W.J.M., Leeuwenberg, H.F.M., Willems, H.M., Capel, P.J.A. & Koene, R.A.P. in *Leukocyte Typing* (ed. Bernard, A. *et al.* 721-724 (Springer-Verlag, Berlin, 1984).
7. Krangel, M.S. *et al. Proc. natn. Acad. Sci. USA* **84**, 3817-3821 (1987).
8. Band, H. *et al. Science* **238**, 682-684 (1987).
9. Wu, Y.-J. *et al. J. Immunol.*, (submitted).
10. Jitsukawa, S., Faure, F., Lipinski, M., Triebel, F. & Hercend, T. *J. exp. Med.* **166**, 1192-1197 (1987).
11. Forster, A., Huck, S., Ghanem, N., Lefranc, M.-P. & Rabbitt, T.H. *EMBO J.* **6**, 1945-1950 (1987).
12. Samelson, L.E., Harford, J., Schwartz, R.H. & Klausner, R.D. *Proc. natn. Acad. Sci. USA* **82**, 1969-1973 (1985).

Monoclonals and more

The unique specificity of monoclonal antibodies has made them handy tools in immunology, molecular genetics and cancer research. This week's review covers a spectrum of monoclonals, ELISAs and separation equipment.

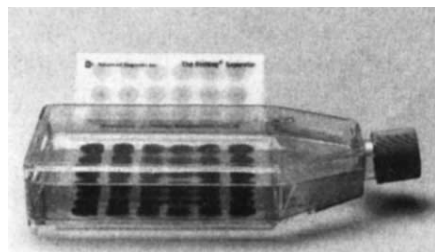
A CELL harvester that offers one-step 96-well harvesting is new from Inotech Biosystems International (Reader Service No. 101). Inotech says its **Cell Harvester System** can harvest a complete 96-well plate in two minutes or less, including washing and cutting filter discs. The discs are automatically deposited into scintillation vials, so no manual handling of discs



Inotech Biosystems cell harvester for 96 well plates

is required. The system is based on one tube per well, a design that eliminates well-to-well carryover contamination. The tubes are used for concurrent washing and harvesting, and their short 2 inch length minimizes background from build-up within the tubes over time. The complete system includes a 96-well harvester, a distributor unit, a cutting/transfer plate, scintillation vial racks and a plate storage stand.

Advanced Magnetics, Inc. sells superparamagnetic particles used for the **separation of cells** and subcellular materials (Reader Service No. 102). The BioMag particles consist of a superparamagnetic core coated to provide amine or carboxyl functional groups. Antibodies, binding proteins, ligands and nucleic acid sequences can be covalently linked to the particles and still maintain their biological activity. The particles are strongly attracted to magnetic fields but do not themselves become permanently magnetized, so targeted cells can be separated from



Advanced Magnetics Bio Mag separator

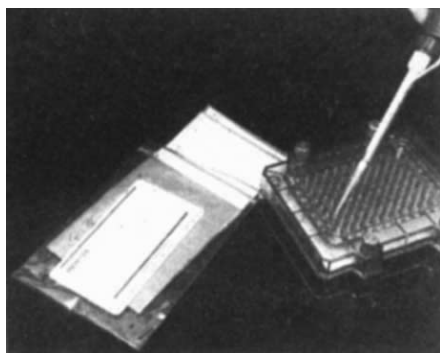
solution using permanent magnets. The targeted cells are held in the magnetic field, while other particles, non-targeted cells or cell debris can be decanted or aspirated off. BioMag products are also available already coated with specific binding proteins such as specific immunoglobulins, protein A, biotin, avidin and sheep anti-fluorescein.

For cellular localization and related techniques, Cellon Sarl now offers a range of **fluorescently-labelled microspheres** (Reader Service No. 103). The latex microspheres are available in 0.5, 0.7 and 0.9 μ m diameters, labelled with red, green or blue high intensity fluorescent labels, or unlabelled. The microspheres are resistant to quenching or leaching, so conventional histochemical stains may be used on the same samples, allowing cell morphology to be related directly to antigen localization. Two types of surface properties are available: the MX microspheres allow the attachment of proteins by simply mixing in physiological saline, and the CX microspheres have free carboxyl groups for the covalent bonding of avidin, lectins or antigens.

Membranes and columns

Hydroxyapatite has been recognized for years as a material that can be used to separate samples without destroying their biological activity. Now Toa Nenryo Kogyo K.K. (Tonen) has developed a line of columns for HPLC based on **spherical hydroxyapatite** (Reader Service No. 104). Tonen says the columns are especially suited for the separation of antibodies, and that IgM and all four subclasses of IgG can be resolved using the columns. Tonen sells a 7.5 $\times$ 100 mm high performance column with 2.2 μ m particles for ¥115,000, a 7.5 $\times$ 100 mm standard analytical column with 5 μ m particles for ¥95,000, and a 21.5 $\times$ 100 mm semi-preparatory column with 5 μ m particles for ¥395,000.

A biochemically reactive **dot binding membrane** that covalently binds to the amino group of proteins is sold by Memtek (Reader Service No. 105). Memtek says the Memtest membranes can be used to replace nitrocellulose, nylon, and diazotized papers in ligand binding experiments, immunoassays and autoradiography. No blocking steps are required, and Memtest membranes can be bound with ligand in concentrations as low



Memtek dot binding membrane

as $1 \mu\text{g ml}^{-1}$ and be stored for months prior to use. The Memtest membranes come in 4×5.25 in. cut sheets for \$68 (US) per pack of ten, and in 5.25 in. $\times$ 10 ft. rolls for \$135 (US) each.

Abundant antibodies

Biotx has five affinity-purified **polyclonal antibodies for cancer research** that it says are monospecific and offer a specificity similar to that of monoclonal antibodies (Reader Service No. 106). Antibodies are available to detect products of the Ha-*ras* pan-specific *ras* (Ha-Ki-N-*ras*), *myc* and *fos* oncogenes, and alpha transforming growth factor. Each antibody comes in a 5 ml vial containing $100 \mu\text{g}$ of lyophilized antibody, with a guaranteed shelf life of one year under refrigeration, and costs \$120–150 (US).

Koch-Light Ltd has introduced for sale the **monoclonal antibody to human interleukin-1 (IL-1)** developed by Thomas Luger at the University of Vienna (Reader Service No. 107). The capability to block the activity of human IL-1 will allow *in vitro* studies to establish the differences in activity between the alpha and beta forms, and to determine whether the biological effects of IL-1 are regulated by a family of related molecules or to a single entity. Koch-Light says the new IgM antibody is suitable for Western blotting, immunoprecipitation, immunopurification, ELISA, RIA and the neutralization of bioassays. The 90 per cent pure antibody is supplied sterile in phosphate-buffered saline without azide, and costs £230 (UK) for 1 mg.

A range of 22 **CD-specific monoclonal antibodies** is available from Janssen (Reader Service No. 108). Each antibody is offered unconjugated in a 1 ml presentation containing 0.2 mg ml^{-1} of the antibody and 10 mg BSA in PBS. Janssen also sells the antibodies in sets, where each antibody is present in 200 μl . Set 1 contains antibodies specific to CD3, CD4 and CD8 and costs £155 (UK), and Set 2 contains CD2, CD3, CD5 and CD8 monoclonals for £215 (UK).

Cambridge Research Biochemicals is developing a range of monoclonal antibodies specific to **oncoproteins** (Reader

Service No. 109). Antibodies scheduled for the range include those for the L-*myc*, *ras*, *src*, *erb-B*, *mas*, *abl*, *neu*, *INT-1*, *INT-2* and *ets* oncoproteins. Pan-specific antibodies for both *myc* and *fos* oncoprotein families are currently available, as are antibodies to N-*myc*, C-*myc*, C-*fms*, C-*fos*, C-*mos*, N-*myb*, C-*myb* and epidermal growth factor. Cambridge Research Biochemicals raises the antibodies to synthetic peptides made using the Fmoc-polyamide mode of solid phase peptide synthesis, and tests them both against synthetic peptide and against native protein released from cells expressing particular oncoproteins by controlled lysis. The antibody preparations cost \$250 (US) for $250 \mu\text{g}$ of purified immunoglobulin and are fully characterized by ELISA, immunoblotting and immunoprecipitation.

Biomedical Technologies, Inc. has developed a rabbit **antibody to fluorescein** that can be used as a "bridge" for labelling samples with another immunohistochemical reagent, such as rhodamine (Reader Service No. 110). The company says the antibody allows specimens prepared for light microscopy to be used for electron microscopy as well. For example, cells labelled with fluorescein phalloidin for fluorescence microscopy can be labelled with a rhodamine tag for electron microscopy by reacting the cells with anti-fluorescein and then reacting them with a rhodamine-labelled anti-rabbit IgG. Gold-labelled or peroxidase-labelled antibodies can also be used for electron or light microscopy. The company sells anti-fluorescein in 2 ml aliquots for \$90 (US).

ELISAs

The level of released CD8 antigen in body fluids has recently been shown to be an indicator of suppressor/cytotoxic cell activity. An enzyme immunoassay for the **quantitation of soluble CD8 antigen** in serum, plasma or culture supernatants is now available from T Cell Sciences (Reader Service No. 111). The Cellfree test kit is an ELISA assay that can be performed on fresh or frozen samples, providing the opportunity for retrospective analysis of stored samples. The \$485 (US) kit is for research only, and includes monoclonal antibodies, standards, diluents and buffers.

Ultraclean, frosted $1/4$ in. and $5/16$ in. **polystyrene beads for ELISA and RIA** applications are available from Interfacial Dynamics (Reader Service No. 112). The beads are available in quantities of 100 to 100,000 or larger, and have been produced without the additives normally used in the fabrication of plastics, such as mould release agents. The area per bead has been determined for each batch of beads from the adsorption of ultrapure proteins, and this data is supplied as part



Interfacial Dynamics ELISA beads

of the certification for each shipment. For further assurance, Interfacial Dynamics certifies the reproducibility of the physicochemical properties of its macrobeads.

Cistron Biotechnology has introduced an ELISA kit for the measurement of human **interleukin-1B (IL-1B)** (Reader Service No. 113). The company says the \$425 (US) kit is ten times more sensitive and eight times faster than the existing RIA, and is more reproducible than bioassays. The kit performs 96 tests, and consists of a 96-well plate coated with monoclonal antibody to human IL-1B, plus IL-1B rabbit antiserum, anti-rabbit IgG conjugated with horseradish peroxidase, IL-1B standards and associated buffers. Cistron reports a sensitivity of 20 pg ml^{-1} for the kit.

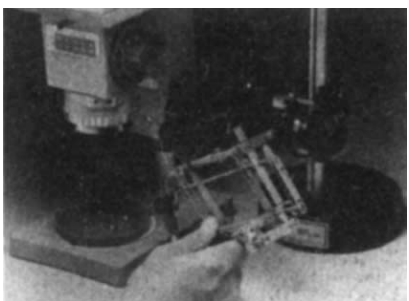
Setting it straight

A review of the animal activity monitor from HVS Image Analysing Ltd, published in the 2 July issue of *Nature* (328, p.93), contained a misleading typographical error. The monitor uses an overhead TV camera, not a UV camera, to record the movements of research animals.

These notes are compiled by Carol Ezzell from information provided by the manufacturer. To obtain further details about these products, use the reader service card bound inside the journal. Prices quoted are sometimes nominal, and apply only within the country indicated.

ADVERTISEMENT

SINGER INSTRUMENTS



Singer Mk1 Micromanipulator

Perfect control in oocyte injection, microdissection, cell aspiration and other fine work under the stereo microscope.

Tel: (0984) 40226. Telex: 337492 Comcab G.

Defence skills for civil R&D?

Richard Pearson

With the pressures on research spending in the United Kingdom mounting, covetous eyes are again turning towards the country's sizeable defence budget.

In the United Kingdom over half of the government's expenditure on research and development (R&D) goes on military work. Overall the UK allocates a higher share of its gross domestic product (GDP) to defence R&D than any other major western government apart from the United States. Critics of this high level of defence R&D expenditure argue that it is diverting scarce funds away from civil areas and that is an important reason for Britain's poor industrial performance. Supporters point to nearly £6,000 million export sales and its role as a major job creator.

UK defence R&D cost £2,400 million in 1986, just over half of all that government spent on R&D and roughly equal to the total R&D bill of the private sector. Military expenditure is clearly high and over half of it is spent in the information technology sector involving electronics, aerospace and related products. This is also a sector experiencing persistent skill shortages, so to what extent is military expenditure 'crowding out' civil needs and is it a contributor to skill shortages? Is there technology transfer as people move from military activities to the civil sector? And who would benefit were it to be cut back?

The information technology industry is itself a major industry in the United Kingdom with sales in excess of £8,000 million per annum. There are 200,000 professional information technology staff at graduate level and above, 70,000 of them employed in the electronics and computer industries, a further 40,000 in the software houses and consultancies, while the remaining 90,000 are employed by users of information technology in all sectors of the economy, including the public sector. One-third have electronics-based skills, two-thirds computing-based skills.

Where then does the defence sector fit? Unfortunately, data on defence-related employment are scarce, a problem compounded by defence activity being spread across many sectors of the economy, involving not just the big electronics and aerospace companies, but also software and engineering companies and a myriad of suppliers. The universities also receive well over £10 million from the Ministry of Defence each year as well as more from the defence contractors. The government's own research establishments employ over 20,000 staff on defence work. MoD R&D expenditure in the private

sector involves another 40,000 staff, of whom up to one-third are graduate-level scientists and engineers, the majority with information technology skills. This suggests there could be as many as 20,000 information technology professionals working on defence R&D, up to 10% of the national total. Data for the United States which are more detailed show that about 15% of engineers work on defence-related activities, a figure not incompatible with that for the United Kingdom given the higher levels of defence expendi-



ture in the United States. Clearly defence R&D is not a dominant employer of information technology skills.

However, analysis at an occupational level shows a high proportion of defence skills to be electronics rather than computing based so a defence employment of some 20,000 out of a UK total of 60,000 electronics engineers becomes rather more significant. Given that there are shortages of these skills and defence contractors usually offer attractive work and associated working conditions, they share rather fewer of the shortages than do some of the less competitive employers in the civil sector. Not all defence employers are, however, well placed in the labour market, as many Ministry of Defence (MoD) research establishments are seriously understaffed and are losing many people to much better paid jobs with the private sector defence contractors. This is giving major problems to the MoD in terms of its procurement programme.

Focusing on graduate recruitment shows that, of the ten largest recruiters of information technology graduates, six have heavy defence interests. For two, the intake for defence-related work accounts for more than half their graduate intake, in one case totalling well over 300 vacancies for information technology

graduates. It is clear therefore that the defence sector takes a significant proportion of the annual output of some 8,000 graduates from information technology departments in higher education.

Another area of concern is the apparent low level of technology transfer from UK defence R&D to civil applications. This can come through many routes, one of which is staff mobility and a number of companies are now actively encouraging research staff to move out from the labs into production areas, taking the latest technology with them and aiding the integration of research and production. Many operating divisions have yet, however, to see this happen to any significant degree. Wastage rates from the electronics companies are usually well under 10 per cent per annum, suggesting only limited inter-company flows; indeed many of these flows are within rather than from the defence community. Two barriers inhibit this flow: first, research staff rarely want to change their type of work while the second concerns the more general and worsening problem of staff relocation. The biggest flows, between types of work, companies and sectors, are at present from the software houses.

It is clear that while the defence sector does not take a massive share of information technology skills it does account for a significant proportion of scarce electronics and software engineering skills.

What then are the likely implications of a cutback in defence expenditure; would it release skills needed by the civil sector? The answer is not an unqualified yes. More scarce new information technology graduates would become available to the civil sector but perhaps more importantly also to higher education which is suffering particularly badly in its ability to attract both teaching and research staff. At the experienced level there is unlikely to be an easy switch across to the civil sector as the short-term skills required by the two sectors are very different, the civil sector having a much stronger emphasis on computing skills. The main beneficiaries are again likely to be the universities and polytechnics. As such, any cutback in defence expenditure is likely to benefit the labour market more in the long term than the short term. □

Richard Pearson is at the Institute of Manpower Studies, Mantell Building, University of Sussex, Brighton BN1 9RF, UK.